

# EViews 3.1

## User's Guide



QUANTITATIVE  
MICRO SOFTWARE

# EViews 3.1 User's Guide

3rd Edition

Copyright © 1994–1999 Quantitative Micro Software, LLC

All Rights Reserved

Printed in the United States of America

This software product, including program code and manual, is copyrighted, and all rights are reserved by Quantitative Micro Software, LLC. The distribution and sale of this product are intended for the use of the original purchaser only. Except as permitted under the United States Copyright Act of 1976, no part of this product may be reproduced or distributed in any form or by any means, or stored in a data base or retrieval system, without the prior written permission of Quantitative Micro Software.

## Disclaimer

The authors and Quantitative Micro Software assume no responsibility for any errors that may appear in this manual or the EViews program. The user assumes all responsibility for the selection of the program to achieve intended results, and for the installation, use, and results obtained from the program.

## Trademarks

Windows, Windows 95, Windows 98, Windows NT, and Microsoft Excel are trademarks of Microsoft Corporation. PostScript is a trademark of Adobe Corporation. All other product names mentioned in this manual may be trademarks or registered trademarks of their respective companies.

Quantitative Micro Software, LLC

4521 Campus Drive, PMB 336, Irvine CA, 92612-2699

Telephone: (949) 856-3368

Fax: (949) 856-2044

e-mail: [sales@eviews.com](mailto:sales@eviews.com)

web: [www.eviews.com](http://www.eviews.com)

# Table of Contents

---

PREFACE .....	1
PART I. EViews FUNDAMENTALS .....	3
CHAPTER 1. INTRODUCTION .....	5
What Is EViews? .....	5
Installing and Running EViews .....	5
Windows Basics .....	6
Closing EViews .....	13
Where To Go For Help .....	13
CHAPTER 2. A DEMONSTRATION .....	15
Creating a Workfile and Importing Data .....	15
Verifying the Data .....	18
Examining the Data .....	20
Estimating a Regression Model .....	22
Specification and Hypothesis Tests .....	24
Modifying the Equation .....	26
Forecasting from an Estimated Equation .....	27
Additional Issues .....	31
CHAPTER 3. EViews BASICS .....	33
Workfile Basics .....	33
Object Basics .....	40
Commands .....	52
CHAPTER 4. BASIC DATA HANDLING .....	53
Samples .....	58
Importing Data .....	62
Exporting Data .....	68
Commands .....	70
Addendum: Reading ASCII Files .....	70
Addendum: Matrix Object Reading and Writing .....	78
CHAPTER 5. WORKING WITH DATA .....	81

Using Expressions . . . . .	81
Working with Series . . . . .	88
Working with Auto-series . . . . .	92
Working with Groups of Series . . . . .	96
Accessing Individual Elements . . . . .	97
Working with Scalars . . . . .	98
<b>CHAPTER 6. EViews DATABASES . . . . .</b>	<b>101</b>
An Overview . . . . .	101
Database Basics . . . . .	102
Working with Objects in Databases . . . . .	105
Database Auto-Series . . . . .	116
The Database Registry . . . . .	118
Querying the Database . . . . .	120
Maintaining the Database . . . . .	127
Foreign Databases . . . . .	129
Connecting to Remote Databases . . . . .	130
Commands . . . . .	138
<b>PART II. BASIC DATA ANALYSIS . . . . .</b>	<b>141</b>
<b>CHAPTER 7. SERIES . . . . .</b>	<b>143</b>
Series Views . . . . .	143
Series Procedures . . . . .	160
Commands . . . . .	170
<b>CHAPTER 8. GROUPS . . . . .</b>	<b>173</b>
Views from a Group Window . . . . .	173
Group Members . . . . .	173
Spreadsheet . . . . .	173
Dated Data Table . . . . .	174
Graphs . . . . .	182
Descriptive Stats . . . . .	187
N-Way Tabulation . . . . .	188
Correlations, Covariances, and Correlogram . . . . .	192
Cross Correlations . . . . .	192
Cointegration Test . . . . .	193
Granger Causality . . . . .	193

---

Label .....	194
Group Procedures .....	194
Commands .....	195
<b>CHAPTER 9. STATISTICAL GRAPHS USING SERIES AND GROUPS .....</b>	<b>197</b>
Distribution Graphs of Series .....	197
Scatter Diagrams with Fit Lines .....	205
Commands .....	213
<b>CHAPTER 10. GRAPHS, TABLES, AND TEXT OBJECTS .....</b>	<b>215</b>
Graphs .....	215
Multiple Graph Objects .....	222
Printing Graphs .....	223
Copying Graphs to Other Windows Programs .....	224
Tables .....	224
Copying Tables to Other Windows Programs .....	225
Text Objects .....	226
Commands .....	226
<b>PART III. BASIC SINGLE EQUATION ANALYSIS .....</b>	<b>229</b>
<b>CHAPTER 11. BASIC REGRESSION .....</b>	<b>231</b>
Equation Objects .....	231
Specifying an Equation in EViews .....	232
Estimating an Equation in EViews .....	235
Equation Output .....	237
Working with Equations .....	243
Estimation Problems .....	247
Commands .....	248
<b>CHAPTER 12. ADDITIONAL REGRESSION METHODS .....</b>	<b>249</b>
Weighted Least Squares .....	249
Heteroskedasticity and Autocorrelation Consistent Covariances .....	251
Two-Stage Least Squares .....	253
Nonlinear Least Squares .....	260
Generalized Method of Moments (GMM) .....	267
Commands .....	271
<b>CHAPTER 13. TIME SERIES REGRESSION .....</b>	<b>273</b>

Serial Correlation Theory . . . . .	273
Testing for Serial Correlation . . . . .	274
Estimating AR Models . . . . .	277
ARIMA Theory . . . . .	281
Estimating ARIMA Models . . . . .	283
Diagnostic Evaluation . . . . .	292
Polynomial Distributed Lags (PDLs) . . . . .	292
Nonstationary Time Series . . . . .	297
Unit Root Tests . . . . .	298
Commands . . . . .	303
<b>CHAPTER 14. FORECASTING FROM AN EQUATION . . . . .</b>	<b>305</b>
Forecasting from Equations in EViews . . . . .	305
Illustration . . . . .	308
Forecast Basics . . . . .	312
Forecasts with Lagged Dependent Variables . . . . .	318
Forecasting with ARMA errors . . . . .	320
Forecasting Equations with Formulas . . . . .	323
Forecasting with Nonlinear and PDL Specifications . . . . .	327
Commands . . . . .	328
<b>CHAPTER 15. SPECIFICATION AND DIAGNOSTIC TESTS . . . . .</b>	<b>329</b>
Background . . . . .	329
Types of Tests . . . . .	329
Residual Tests . . . . .	337
Specification and Stability Tests . . . . .	341
Applications . . . . .	350
Commands . . . . .	355
<b>PART IV. ADVANCED SINGLE EQUATION ANALYSIS . . . . .</b>	<b>357</b>
<b>CHAPTER 16. ARCH AND GARCH ESTIMATION . . . . .</b>	<b>359</b>
The ARCH Specification . . . . .	359
Estimating ARCH models in EViews . . . . .	361
Working with ARCH Models . . . . .	365
Asymmetric ARCH Models . . . . .	368
The Component ARCH Model . . . . .	373
Estimating and Interpreting Models in EViews . . . . .	374

---

Examples . . . . .	376
Commands . . . . .	379
Technical Details . . . . .	379
<b>CHAPTER 17. DISCRETE AND LIMITED DEPENDENT VARIABLE MODELS . . . . .</b>	<b>381</b>
Binary Dependent Variable Models . . . . .	381
Estimating Binary Models in EViews . . . . .	383
Procedures for Binary Equations . . . . .	393
Ordered Dependent Variable Models . . . . .	399
Estimating Ordered Models in EViews . . . . .	400
Views of Ordered Equations . . . . .	403
Procedures for Ordered Equations . . . . .	404
Censored Regression Models . . . . .	405
Estimating Censored Models in EViews . . . . .	406
Procedures for Censored Equations . . . . .	410
Truncated Regression Models . . . . .	415
Procedures for Truncated Equations . . . . .	416
Count Models . . . . .	418
Views of Count Models . . . . .	422
Procedures for Count Models . . . . .	423
Demonstrations . . . . .	423
Commands . . . . .	427
Technical Notes . . . . .	427
<b>CHAPTER 18. THE LOG LIKELIHOOD (LOGL) OBJECT . . . . .</b>	<b>431</b>
Overview . . . . .	431
Specification . . . . .	433
Estimation . . . . .	439
LogL Views . . . . .	440
LogL Procs . . . . .	441
Troubleshooting . . . . .	443
Limitations . . . . .	444
Examples . . . . .	445
<b>PART V. MULTIPLE EQUATION ANALYSIS . . . . .</b>	<b>453</b>
<b>CHAPTER 19. SYSTEM ESTIMATION . . . . .</b>	<b>455</b>
Background . . . . .	455

System Estimation Methods . . . . .	456
How to Create and Specify a System . . . . .	458
Working With Systems . . . . .	463
Commands . . . . .	468
Technical Discussion . . . . .	468
<b>CHAPTER 20. VECTOR AUTOREGRESSION AND ERROR CORRECTION MODELS . . . . .</b>	<b>477</b>
Vector Autoregression Theory . . . . .	477
Estimating a VAR in EViews . . . . .	478
VAR Estimation Output . . . . .	479
Working with a VAR . . . . .	480
Vector Error Correction and Cointegration Theory . . . . .	487
Testing for Cointegration . . . . .	488
Estimating a VEC in EViews . . . . .	495
Working with a VEC . . . . .	496
Forecasting from a VAR or VEC . . . . .	497
Illustration . . . . .	497
Technical Discussion . . . . .	502
Commands . . . . .	504
<b>CHAPTER 21. STATE SPACE MODELS AND THE KALMAN FILTER . . . . .</b>	<b>505</b>
State Space Models . . . . .	505
Specifying State Space Models . . . . .	506
Estimating State Space Models . . . . .	514
Working With the State Space . . . . .	517
Simulation Example . . . . .	519
Commands . . . . .	521
Technical Discussion . . . . .	522
<b>CHAPTER 22. POOLED TIME SERIES, CROSS-SECTION DATA . . . . .</b>	<b>525</b>
Creating a Workfile For Pooled Data . . . . .	525
The Pool Object . . . . .	525
Importing Pooled Data . . . . .	527
Exporting Pooled Data . . . . .	533
Working with Pooled Data . . . . .	533
Pooled Estimation . . . . .	536
Commands . . . . .	545
Technical Discussion . . . . .	545



---

<b>CHAPTER 23. MODELS</b>	<b>551</b>
Creating a Model	551
Specifying a Model	551
Equations in a Model	552
Statements in a Model	555
Solving Models	558
Working with Models	561
Commands	564
<b>APPENDIX A. MATHEMATICAL OPERATORS AND FUNCTIONS</b>	<b>565</b>
Operators	565
Date and Observation Functions	565
Basic Mathematical Functions	566
Time Series Functions	567
Descriptive Statistics	568
Additional and Special Functions	569
Trigonometric Functions	572
Statistical Distribution Functions	572
Single Equation Regression Statistics	576
Systems of Equations Regression Statistics	577
<b>APPENDIX B. GLOBAL OPTIONS</b>	<b>579</b>
Setting Global Options	579
<b>APPENDIX C. DATE FORMATS</b>	<b>583</b>
Date Specification	583
Implicit Dating	584
Special Date Functions	585
Examples	585
<b>APPENDIX D. WILDCARDS</b>	<b>587</b>
Wildcard Expressions	587
Using Wildcard Expressions	587
Resolving Ambiguities	588
Wildcard versus Pool Identifier	589
<b>APPENDIX E. ESTIMATION ALGORITHMS AND OPTIONS</b>	<b>591</b>
Nonlinear Optimization Algorithms	591
Setting Estimation Options	594

Nonlinear Equation Solution Methods .....597

What Does EViews Use? .....598

APPENDIX F. INFORMATION CRITERIA ..... 599

    Definitions .....599

    Using Information Criteria as a Guide to Model Selection .....600

REFERENCES .....601

INDEX .....607

# Preface

---

This manual describes the interactive use of EViews, a program for statistical and econometric analysis, and forecasting. For details on the EViews command language, as well as a description of the programming and matrix languages, we refer you to the companion volume—the EViews *Command and Programming Reference*.

The manual is divided into five parts:

- Part I, “EViews Fundamentals”—introduces you to the basics of using EViews. In addition to a discussion of basic Windows operations, we explain how to use EViews to manage your data.
- Part II, “Basic Data Analysis”—describes the use of EViews to perform basic analysis of data and to draw graphs and create tables describing your data.
- Part III, “Basic Single Equation Analysis”—discusses standard regression analysis: ordinary least squares, weighted least squares, two-stage least squares, nonlinear least squares, time series analysis, specification testing and forecasting.
- Part IV, “Advanced Single Equation Analysis”—documents autoregressive conditional heteroskedasticity (ARCH) models, discrete and limited dependent variable models, and user specified likelihood estimation.
- Part V, “Multiple Equation Analysis”—describes estimation and forecasting with systems of equations, vector autoregression and error correction models, state space models, pooled cross-section/time series data, and model solution.

You should not feel a need to read the manual from cover-to-cover in order to use EViews. We recommend, however, that you glance at most of Part I to gain familiarity with the basic concepts and operation of the program. At a minimum, you should look over Chapters 1–3, especially the extended demonstration in Chapter 2.



## Part I. EViews Fundamentals

---

The following six chapters document the fundamentals of working with EViews.

- Chapters 1–3 contain introductory material describing the basics of Windows and EViews. The demo in Chapter 2 guides you through a typical EViews session.
- Chapters 4 and 5 document the basics of working with data. We describe methods of getting your data into EViews, using EViews to manipulate and manage your data, and exporting your data from EViews into spreadsheets, text files and other Windows applications.
- Chapter 6 contains more advanced material, describing the EViews database and other advanced data handling features.

We recommend that you read through most of the material in Part I before beginning serious work with EViews. Chapters 1–3 in particular should be read prior to using the program.



# Chapter 1. Introduction

---

## What Is EViews?

EViews provides sophisticated data analysis, regression, and forecasting tools on Windows-based computers. With EViews you can quickly develop a statistical relation from your data and then use the relation to forecast future values of the data. Areas where EViews can be useful include: scientific data analysis and evaluation, financial analysis, macroeconomic forecasting, simulation, sales forecasting, and cost analysis.

EViews is a new version of a set of tools for manipulating time series data originally developed in the Time Series Processor software for large computers. The immediate predecessor of EViews was MicroTSP, first released in 1981. Though EViews was developed by economists and most of its uses are in economics, there is nothing in its design that limits its usefulness to economic time series. Even quite large cross-section projects can be handled in EViews.

EViews provides convenient visual ways to enter data series from the keyboard or from disk files, to create new series from existing ones, to display and print series, and to carry out statistical analysis of the relationships among series.

EViews takes advantage of the visual features of modern Windows software. You can use your mouse to guide the operation with standard Windows menus and dialogs. Results appear in windows and can be manipulated with standard Windows techniques.

Alternatively, you may use EViews' powerful command and batch processing language. You can enter and edit commands in the command window. You can create and store the commands in programs that document your research project for later execution.

## Installing and Running EViews

Before beginning the installation procedure, you should make certain that your computer satisfies the following requirements:

- A computer running Microsoft Windows 95 (or later) or Windows NT 4.0 (or later).
- At least 16 megabytes of random access memory (RAM). We recommend, however, that for optimal performance you have at least 32 megabytes of RAM.
- A VGA, super VGA, or compatible display.

- A Windows compatible mouse, trackball, or pen pad.
- Installation sizes will vary, but 10 megabytes of available hard disk space will be more than sufficient.

Make certain that you have all of your EViews diskettes handy and that Disk 1 is not write-protected. To verify that the diskette is not protected, hold the diskette face-up with the label reading “EViews” positioned at the top. The hole in the upper right corner should be covered by a plastic tab. If the hole is open, turn the disk over and slide the tab so that the hole is closed.

- If necessary, type WIN to start Windows.
- Once Windows is running, *we strongly recommend that you close all other Windows applications before beginning the installation procedure and that following installation, you reboot the computer prior to first launching EViews.* Other applications may interfere with the installation program.
- Insert the Install Disk (Disk 1) in the appropriate drive (generally A:).
- Click the **File** menu in the Program Manager and click **Run** (Windows 3.1), or click **Start** and then **Run** (Windows 95).
- Type A:SETUP and press ENTER (use another drive letter in place of A:, if appropriate).

The EViews installer will prompt you for each disk. When it has read through all of the disks, it will prompt you to reinsert Disk 1. Once the installation procedure is complete, the installer will inform you that EViews has been successfully installed.

## Windows Basics

We assume that you are familiar with the basics of Windows, but for the benefit of those who are relatively new to Windows, and for those who can use a refresher course, we provide a brief discussion of some useful techniques, concepts, and conventions that we will use in this manual. We urge those who desire more detail to obtain one of the (many) very good books on Windows.

### The Mouse

EViews uses both buttons of the standard Windows mouse. Unless otherwise specified, when we say that you should *click* on an item, we mean a single click of the left mouse-button. *Double-click* means to click the left mouse-button twice in rapid succession. We will often refer to *dragging* with the mouse; this means that you should click and hold the left mouse-button down while moving the mouse.



## Window Control

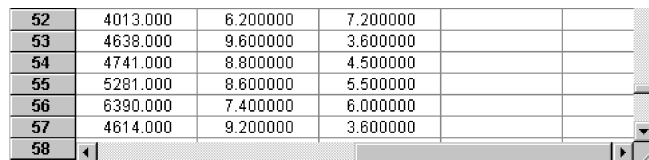
As you work, you may find that you wish to change the size of a window or temporarily move a window out of the way. Alternatively, a window may not be large enough to display all of your output, so that you want to move within the window in order to see relevant items. Windows provides you methods for performing each of these tasks.

### Changing the active window

When working in Windows, you may find that you have a number of open windows on your screen. The *active* (top-most) *window* is easily identified since its title bar will generally differ (in color and/or intensity) from the inactive windows. You can make a window active by clicking anywhere in the window, or by clicking on the word **Window** in the main menu, and selecting the window by clicking on its name.

### Scrolling

Windows provides both horizontal and vertical *scroll bars* so that you can view information which does not fit inside the window (when all of the information in a window fits inside the viewable area, the scroll bars will be hidden).



52	4013.000	6.200000	7.200000		
53	4638.000	9.600000	3.600000		
54	4741.000	8.800000	4.500000		
55	5281.000	8.600000	5.500000		
56	6390.000	7.400000	6.000000		
57	4614.000	9.200000	3.600000		
58					

The scroll box indicates the overall relative position of the window and the data. Here, the vertical scroll box is near the bottom, indicating that the window is showing the lower portion of our data. The size of the box also changes to show you the relative sizes of the amount of data in the window and the amount of data that is offscreen. Here, the current display covers roughly half of the horizontal contents of the window.

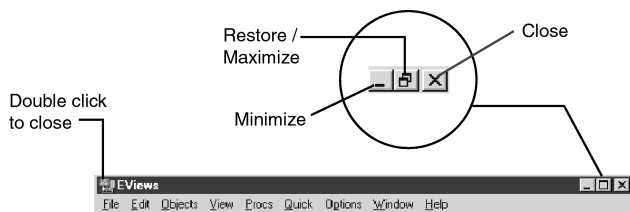
Clicking on the up, down, left, or right scroll arrows on the scroll bar will scroll the display one line in that direction. Clicking on the scroll bar on either side of a scroll box moves the information one screen in that direction.

If you hold down the mouse button while you click on or next to a scroll arrow, you will scroll continuously in the desired direction. To move quickly to any position in the window, drag the scroll box to the desired position.

### Minimize/Maximize/Restore/Close

There may be times that you wish to move EViews out of the way while you work in another Windows program. Or you may wish to make the EViews window as large as possible by using the entire display area.

In the upper right-hand corner of each window, you will see a set of buttons which control the window display:



By clicking on the middle (Restore/Maximize) button, you can toggle between using your entire display area for the window, and using the original window size. *Maximize* (□) uses your entire monitor display for the application window. *Restore* (▢) returns the window to its original size, allowing you to view multiple windows. If you are already using the entire display area for your window, the middle button will display the icon for restoring the window, otherwise it will display the icon for using the full screen area.

You can *minimize* your window by clicking on the minimize button in the upper right-hand corner of the window. To *restore* a program that has been minimized, click on the icon in your task-bar.

Lastly, the *close* button provides you with a convenient method for closing the window. To close all of your open EViews windows, you may also select **Window** in the main menu, and either **Close All**, or **Close All Objects**.

### Moving and Resizing

You can move or change the size of the window (if it is not maximized or minimized). To move your window, simply click on the title bar (the top of your application window) and drag the window to a new location. To resize, simply put the cursor on one of the four sides or corners of the window. The cursor will change to a double arrow. Drag the window to the desired size, then release the mouse button.

### Selecting and Opening Items

To select a single item, you should place the pointer over the item and single click. The item will now be highlighted. If you change your mind, you can change your selection by clicking on a different item, or you can cancel your selection by clicking on an area of the window where there are no items.

You can also select multiple items:

- To select sequential items, click on the first item you want to select, then drag the cursor to the last item you want to select and release the mouse button. All of the items will be

selected. Alternatively, you can click on the first item, then hold down the SHIFT key and click on the last item.

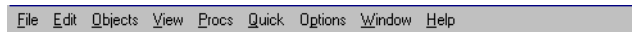
- To select non-sequential items, click on the first item you want to select, then holding the CTRL key, click on each additional item.
- You can also use CTRL-click to “unselect” items which have already been selected. In some cases it may be easier to select a set of sequential items then unselect individual items.

Double clicking on an item will usually open the item. If you have multiple items selected, you can double click anywhere in the highlighted area.

## Menus and Dialogs

Windows commands are accessed via *menus*. Most applications contain their own set of menus, which are located on the menu bar along the top of the application window. There are generally drop-down menus associated with the items in the main menu bar.

For example, the main EViews menu contains:



Selecting **File** from this menu will open a drop-down menu containing additional commands. We will describe the EViews menus in greater detail in the coming sections.

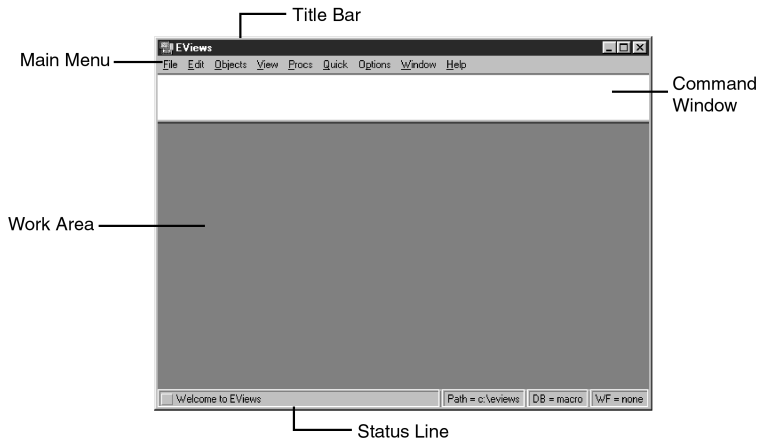
There are a few conventions which Windows uses in its menus that are worth remembering:

- A grayed-out command means the command is not currently available.
- An ellipse (...) following the command means that a dialog box (prompting you for additional input) will appear before the command is executed.
- A right-triangle (▸) means that additional (cascading) menus will appear if you select this item.
- A check mark (✓) indicates that the option listed in the menu is currently in effect. If you select the item again, the option will no longer be in effect and the check mark will be removed. This behavior will be referred to as *toggling*.
- Most menu items contain underlined characters representing keyboard shortcuts. You can use the keyboard shortcuts to the commands by pressing the ALT key, and then the underlined character. For example ALT-F in EViews brings up the **File** drop-down menu.
- If you wish to close a menu without selecting an item, simply click on the menu name, or anywhere outside of the menu. Alternatively, you can press the ESC key.

We will often refer to entering information in *dialogs*. Dialogs are boxes that prompt for additional input when you select certain menu items. For example, when you select the menu item to

run a regression, EViews opens a dialog prompting you for additional information about the specification, and often suggests default values for arguments. You can always tell when a menu item opens a dialog by the ellipses in the drop-down menu entry. **The EViews Window**

If the program is correctly installed, you should see the EViews window when you launch the program. This is what the EViews window looks like:



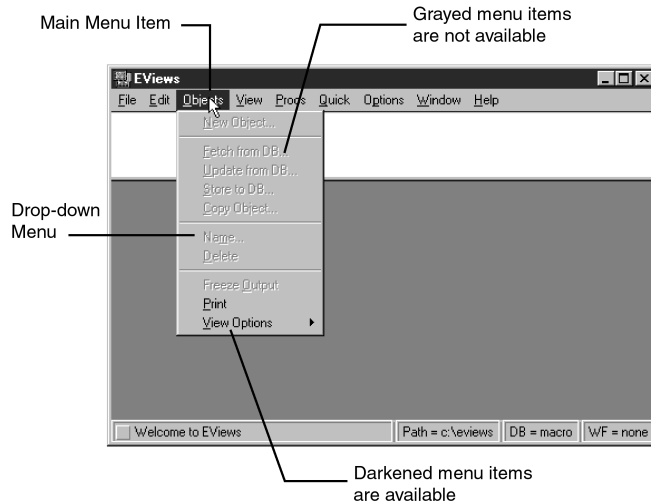
You should familiarize yourself with the following main areas in the EViews Window.

## The Title Bar

The *title bar*, labeled **EViews**, is at the very top of the main window. When EViews is the active program in Windows, the title bar has a color and intensity that differs from the other windows (generally it is darker). When another program is active, the EViews title bar will be lighter. If another program is active, EViews may be made active by clicking anywhere in the EViews window or by using ALT-TAB to cycle between applications until the EViews window is active.

## The Main Menu

Just below the title bar is the *main menu*. If you move the cursor to an entry in the main menu and click on the left mouse button, a *drop-down menu* will appear. Clicking on an entry in the drop-down menu selects the highlighted item.



For example, here we click on the Objects entry in the main menu to reveal a drop-down menu. Notice that some of the items in the drop-down are listed in black and others are in gray. In menus, black items may be executed while the gray items are not available. In this example, you cannot create a **New Object** or **Store** an object, but you can **Print** and **View Options**. We will explain this behavior in our discussion of “The Object Window” on page 45 of Chapter 3.

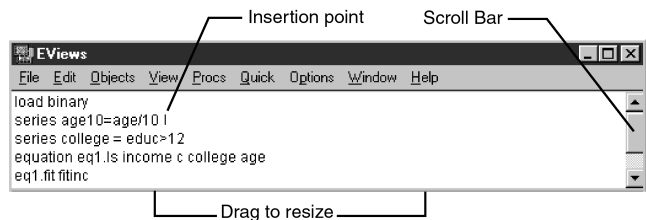
## The Command Window

Below the menu bar is an area called the *command window*. EViews commands may be typed in this window. The command is executed as soon as you hit ENTER.

The vertical bar in the command window is called the *insertion point*. It shows where the letters that you type on the keyboard will be placed. As with standard word processors, if you have typed something in the command

area, you can move the insertion point by pointing to the new location and clicking the mouse. If the insertion point is not visible, it probably means that the command window is not active; simply click anywhere in the command window to tell EViews that you wish to enter commands.

You can move the insertion point to previously executed commands, edit the existing command, and then press ENTER to execute the edited version of the command.



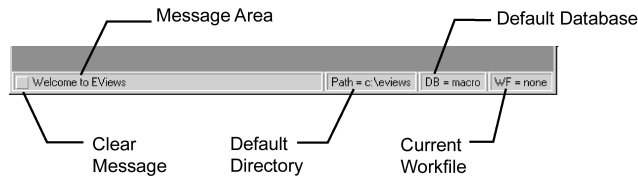
The command window supports Windows cut-and-paste so that you can easily move text between the command window, other EViews text windows, and other Windows programs. The contents of the command area may also be saved directly into a text file for later use: make certain that the command window is active by clicking anywhere in the window, and then select **File/Save As...** from the main menu.

If you have entered more commands than will fit in your command window, EViews turns the window into a standard scrollable window. Simply use the scroll bar or up and down arrows on the right-hand side of the window to see various parts of the list of previously executed commands.

You may find that the default size of the command window is too large or small for your needs. You can resize the command window by placing the cursor at the bottom of the command window, holding down the mouse button and dragging the window up or down. Release the mouse button when the command window is the desired size.

## The Status Line

At the very bottom of the window is a *status line* which is divided into several sections.



The left section will sometimes contain status messages sent to you by EViews. These status messages can be cleared manually by clicking on the box at the far left of the status line. The next section shows the *default directory* that EViews will use to look for data and programs. The last two sections display the names of the default database and workfile. In later chapters, we will show you how to change both defaults.

## The Work Area

The area in the middle of the window is the work area where EViews will display the various object windows that it creates. Think of these windows as similar to the sheets of paper you might place on your desk as you work. The windows will overlap each other with the foremost window being in focus or active. Only the active window has a darkened titlebar.

When a window is partly covered, you can bring it to the top by clicking on its titlebar or on a visible portion of the window. You can also cycle through the displayed windows by pressing the F6 or CTRL-TAB keys.

Alternatively, you may directly select a window by clicking on the **Window** menu item, and selecting the desired name.

You can move a window by clicking on its title bar and dragging the window to a new location. You can change the size of a window by clicking at the lower right corner and dragging the corner to a new location.

## Closing EViews

There are a number of ways to close EViews. You can always select **File/Exit** from the main menu, or you can press ALT-F4. Alternatively, you can click on the close box in the upper right-hand corner of the EViews window, or double click on the EViews icon in the upper left-hand corner of the window. If necessary, EViews will warn you and provide you with the opportunity to save any unsaved work.

## Where To Go For Help

### The EViews Manuals

This *User's Guide* describes how to use EViews to carry out your research. The earlier chapters deal with basic operations, the middle chapters cover basic econometric methods, and the later chapters describe more advanced methods.

Though we have tried to be complete, it is not possible to document every aspect of EViews. There are almost always several ways to do the same thing in EViews, and we cannot describe them all. In fact, one of the strengths of the program is that you will undoubtedly discover alternative, and perhaps more efficient, ways to get your work done.

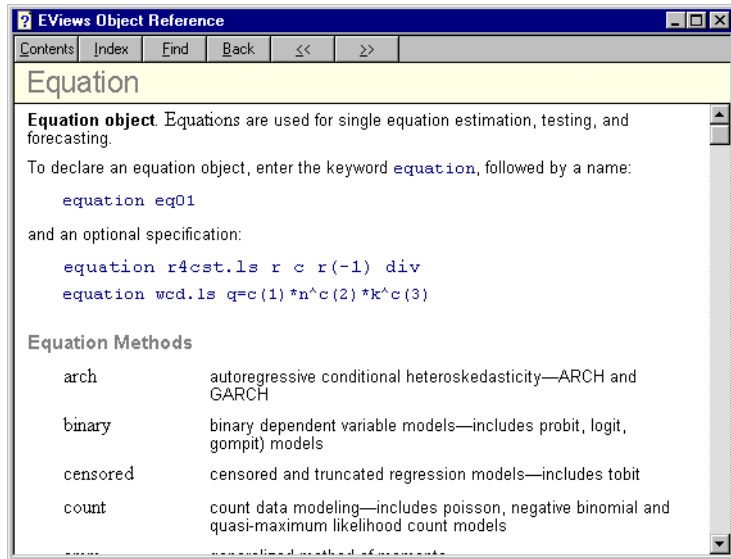
Most of the *User's Guide* explains the visual approach to using EViews. It describes how you can use your mouse to perform operations in EViews. To keep the explanations simple, we do not tell you about alternative ways to get your work done. For example, we will not remind you about the ALT- keyboard alternatives to using the mouse.

When we get to the discussion of the substantive statistical methods available in EViews, we will provide some technical information about the methods, and references to econometrics textbooks and other sources for additional information.

### The Help System

Almost all of the EViews documentation may be viewed from within EViews by using the help system. To access the EViews help system, simply go to the main menu and select **Help**.

Since EViews uses standard Windows Help, the on-line manual is fully searchable and hypertext linked. You can set bookmarks to frequently accessed pages, and annotate the on-line documentation with your own notes.



In addition, the Help system will contain updates to the documentation that were made after the manuals went to press.

## The World Wide Web

In addition to the information provided in the manuals and the help system, we have set up information areas on the Web that you may access using your favorite browser. You can find answers to common questions about installing, using, and getting the most out of EViews.

Another popular area is our Download Section, which contains on-line updates to EViews Version 3.1, sample data and programs, and much more. Your purchase of EViews provides you with much more than the enclosed program and printed documentation. As we make minor changes and revisions to the current version of EViews, we will post them on our web site for you to download. As a valued QMS customer, you are free to download updates to the current version as often as you wish.

So set a bookmark to our site and visit often; the address is: <http://www.eviews.com>.



## Chapter 2. A Demonstration

---

In this chapter, we provide a demonstration of the basic features of EViews. This demo is not meant to be a comprehensive description of the program. A full description of the program begins in Chapter 3.

This demo shows takes you through the following steps:

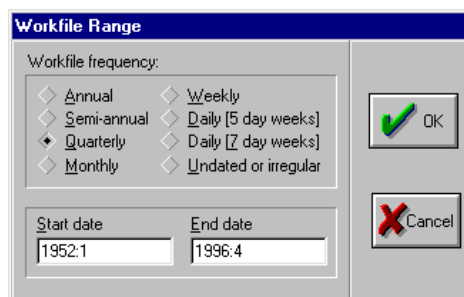
- importing data into EViews from an Excel spreadsheet
- examining the data and performing simple statistical analysis
- using regression analysis to model and forecast a statistical relationship
- performing specification and hypothesis testing
- plotting results

### Creating a Workfile and Importing Data

The first step in the project is to read the data into an EViews workfile.

Before we describe the process of importing data, note that the demonstration data have been included in your EViews directory in both Excel spreadsheet and EViews workfile formats. If you wish to skip the discussion of importing data and go directly to the analysis part of the demonstration, you may load these data by selecting **File/Open/Workfile...** and opening DEMO.WF1.

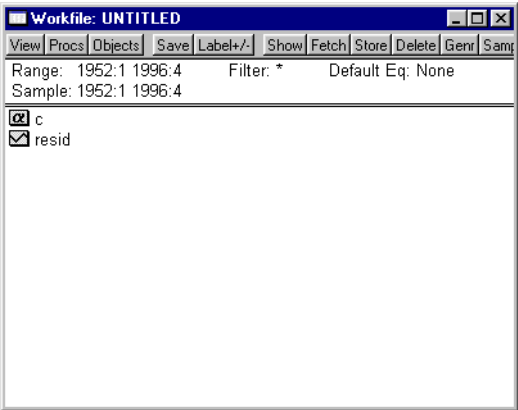
To create a workfile to hold your data, select **File/New/Workfile...**, which opens a dialog box where you will provide information about your data:



For our example, quarterly data are observed from the first quarter of 1952 to the end of 1996. You should set the workfile frequency to quarterly, and specify the start date 1952:1, and the end date 1996:4.

Once you have filled out the dialog, click on the **OK** button. EViews will create an untitled workfile, and will display a workfile window.

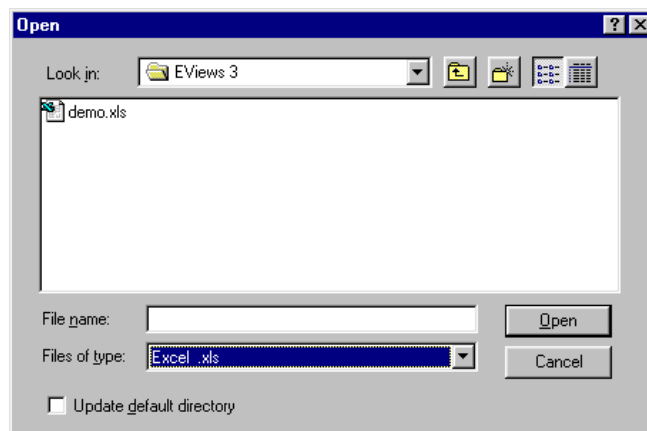
The workfile window is described in detail in Chapter 3, beginning on page 35. For now, notice that the workfile window displays two pairs of dates: one for the range of dates contained in the workfile, and the second for the current workfile sample. Note also that the workfile contains the coefficient vector **C** and the series **RESID**. All EViews workfiles will contain these two objects.



The next step is to import data into the workfile. The data for the four variables used in the analysis have been provided in an Excel file named DEMO.XLS. The data in the DEMO.XLS are arranged with each of the four series in columns, with names in the first row, and dates in the first column:

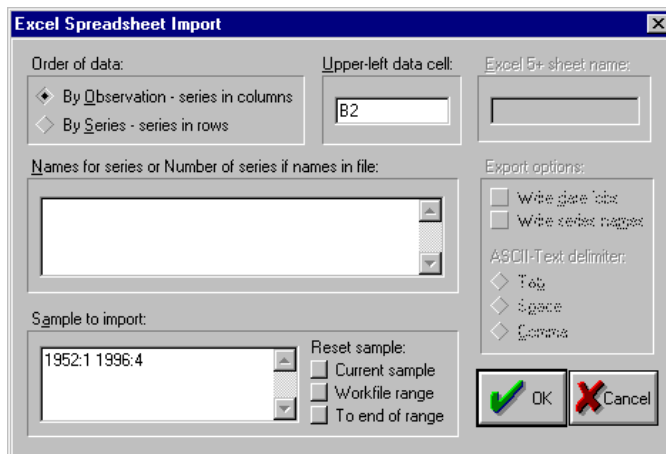
	A	B	C	D	E	F	G
1	OBS	GDP	PR	M1	RS		
2	1952:1	87.875	0.197561	126.537	1.64		
3	1952:2	88.125	0.198167	127.506	1.677667		
4	1952:3	89.625	0.200179	129.385	1.828667		
5	1952:4	92.875	0.201246	128.512	1.923667		
6	1953:1	94.625	0.201052	130.587	2.047333		
7	1953:2	95.55	0.201444	130.341	2.202667		
8	1953:3	95.425	0.202236	131.389	2.021667		
9	1953:4	94.175	0.202723	129.891	1.486333		
10	1954:1	94.075	0.203416	130.173	1.083667		
11	1954:2	94.2	0.203841	131.385	0.814333		
12	1954:3	95.45	0.204291	134.627	0.869667		
13	1954:4	97.36375	0.204374	134.252	1.036333		
14	1955:1	100.725	0.205603	136.413	1.256333		
15	1955:2	102.825	0.206227	136.471	1.614333		
16	1955:3	104.925	0.207762	138.377	1.861333		

To read these data, click on **Procs/Import/Read Text-Lotus-Excel...**, which opens the following file open dialog:

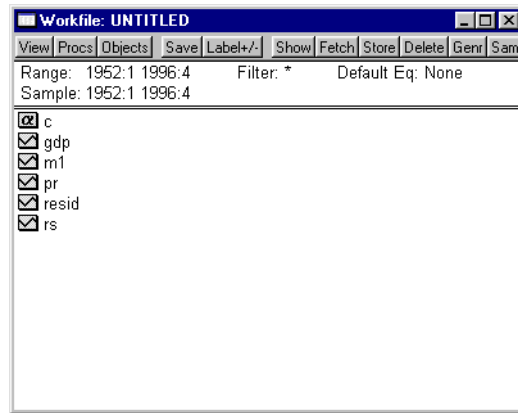


Locate the DEMO.XLS file (it should be in your EViews installation directory) and double click on the file name. You can make finding the file a bit easier by choosing to display **Excel .xls** files from the **Files of type** combo box.

EViews will open the Excel spreadsheet import dialog:



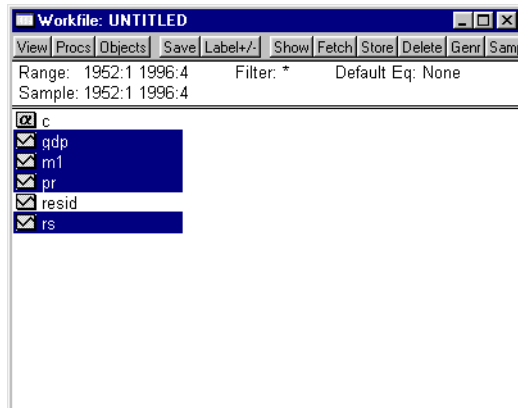
The default settings for order of data, upper-left data cell, and the sample to import should be appropriate for this Excel file. Since the names of the series are in the first row of the Excel file, you can simply enter the number of series (in this case you will want to enter “4”), in the **Names for series or Number of series if name in file** field of the dialog box. Click **OK**, and EViews will import the four series. These series will appear as icons in the workfile window:



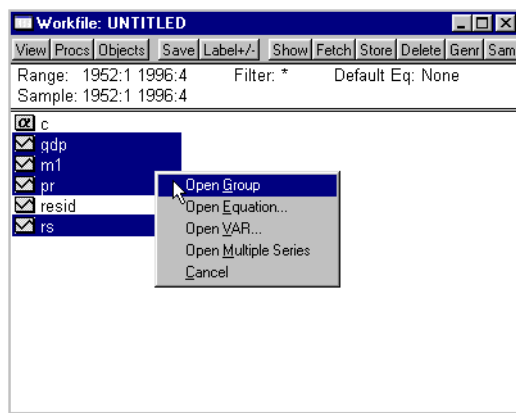
An alternative method of importing the data is to copy-and-paste the data from the Excel spreadsheet directly into EViews. This procedure is described in detail in Chapter 4, “Copying and Pasting” on page 64.

## Verifying the Data

The first thing you should do after importing the data is to verify that the data have been read correctly. We will create a group object that allows us to examine all four series. Click on the name GDP in the workfile window, and then press CTRL and click on M1, PR, and RS. All four of the series should be highlighted:



Now place the cursor anywhere in the highlighted area and double click the left mouse button. EViews will open a popup menu providing you with several options:



Choose **Open Group**. EViews will create an untitled group object containing all four of the series. The default window for the group shows a spreadsheet view of the series:

The screenshot shows the 'Group: UNTITLED' window in spreadsheet view. The window title is 'Group: UNTITLED Workfile: UNTITLED'. The 'Objects' tab is active, showing a table with columns for the series: GDP, PR, M1, and RS. The data is organized into rows for each series, with the first row labeled 'ob1' and subsequent rows labeled with time periods (e.g., 1952:1, 1952:2, etc.).

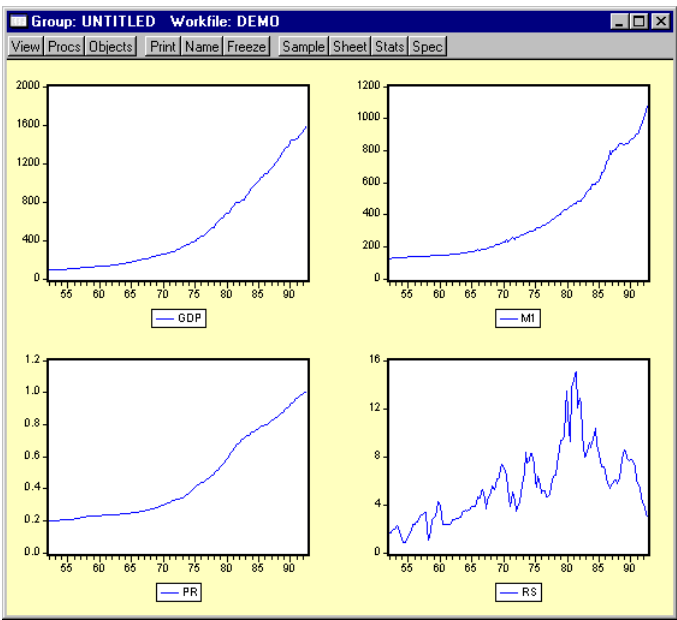
ob1	GDP	PR	M1	RS
1952:1	87.87500	0.197561	126.5370	1.640000
1952:2	88.12500	0.198167	127.5090	1.677667
1952:3	89.62500	0.200179	129.3890	1.828667
1952:4	92.87500	0.201246	128.5120	1.923667
1953:1	94.62500	0.201052	130.5870	2.047333
1953:2	95.55000	0.201444	130.3410	2.202667
1953:3	95.42500	0.202236	131.3890	2.021667
1953:4	94.17500	0.202723	129.9910	1.486333
1954:1	94.07500	0.203416	130.1730	1.063667
1954:2	94.20000	0.203841	131.3890	0.814333
1954:3	95.45000	0.204291	134.6270	0.869667
1954:4	97.36375	0.204374	134.2520	1.036333
1955:1	100.7250	0.205603	136.4130	1.256333
1955:2	100.2000	0.206003	136.4130	1.256333
1955:3	100.2000	0.206003	136.4130	1.256333

You should compare the spreadsheet view with the top of the Excel worksheet to insure that the first part of the data has been read correctly. You can use the scroll bars and scroll arrows on the right side of the window to verify the remainder of the data.

Once you are satisfied that the data are correct, you should save the workfile by clicking the **Save** button in the workfile window. A save dialog will open, prompting you for a workfile name and location. You should enter DEMO2, then click OK. EViews will save the workfile in the specified directory with the name DEMO2.WF1. A saved workfile can be opened later by selecting **File/Open/Workfile...** from the main menu.

## Examining the Data

We can use basic EViews tools to examine the data in a variety of ways. For example, if you select **View/Multiple Graphs/Line** from the group object toolbar, EViews displays line graphs of each of the series.



You can select **View/Descriptive Stats/Individual Samples** to compute descriptive statistics for each of the series:

The screenshot displays the EViews software interface with a group object named 'UNTITLED' and a workfile named 'UNTITLED'. The 'View' menu is open, and the 'Descriptive Stats/Individual Samples' option is selected. A table of descriptive statistics is shown for the four series: GDP, M1, PR, and RS. The table includes statistics such as Mean, Median, Maximum, Minimum, Std. Dev., Skewness, Kurtosis, Jarque-Bera, and Probability, along with the number of observations for each series.

	GDP	M1	PR	RS
Mean	632.4190	445.0064	0.514106	5.412928
Median	374.3000	298.3990	0.383802	5.057500
Maximum	1948.225	1219.420	1.110511	15.08733
Minimum	87.87500	126.5370	0.197561	0.814333
Std. Dev.	564.2441	344.8315	0.303483	2.909939
Skewness	0.843527	0.995000	0.591063	0.984037
Kurtosis	2.331980	2.672168	1.819076	4.027384
Jarque-Bera	24.69301	30.50682	20.94002	36.96626
Probability	0.000004	0.000000	0.000028	0.000000
Observations	180	180	180	180

or click on **View/Correlations** to display the correlation matrix of the four series:

Group: UNTITLED    Workfile: UNTITLED

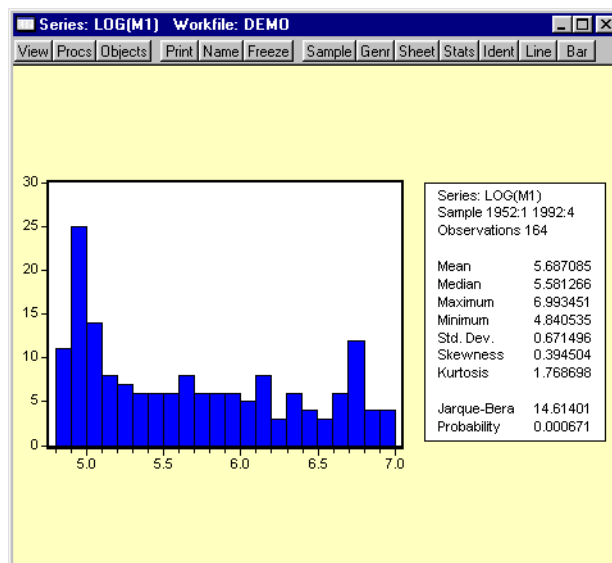
View   Procs   Objects   Print   Name   Freeze   Sample   Sheet   Stats   Spec

Correlation Matrix

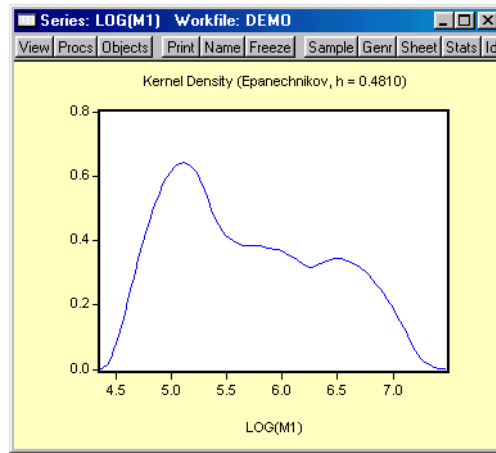
	GDP	M1	PR	RS
GDP	1.000000	0.995197	0.992475	0.333494
M1	0.995197	1.000000	0.980402	0.270059
PR	0.992475	0.980402	1.000000	0.412471
RS	0.333494	0.270059	0.412471	1.000000

We can also examine characteristics of the individual series. Since our regression analysis below will be expressed in logarithms, we will work the log of M1. Select **Quick/Show...** then enter  $\log(m1)$ , and click OK. EViews will open a series window for LOG(M1).

Now select **View/Descriptive Statistics/Histogram and Stats** from the series toolbar to display the descriptive statistics for LOG(M1):



We can construct a smoothed version of the histogram by selecting **View/Distribution Graphs/Kernel Density...** and clicking on OK to accept the default options:



## Estimating a Regression Model

We now estimate a regression model for M1 using data over the period from 1952:1–1992:4 and use this estimated regression to construct forecasts over the period 1993:1–2003:4. The model specification is

$$\log(M1_t) = \beta_1 + \beta_2 \log GDP_t + \beta_3 RS_t + \beta_4 \Delta \log PR_t + \epsilon_t \quad (2.1)$$

where  $\log(M1)$  is the logarithm of the money supply,  $\log(GDP)$  is the log of income,  $RS$  is the short term interest rate, and  $\Delta \log(PR)$  is the log first difference of the price level (the approximate rate of inflation).

To estimate the model, we will create an equation object. Select **Quick** from the main menu and choose **Estimate Equation...** to open the estimation dialog. Enter the following equation specification:

The "Equation Specification" dialog box is shown. It has a title bar "Equation Specification" and a close button. The "Equation Specification:" section contains a text box with the text "log(M1) c log(GDP) RS dlog(PR)". Below this is a scrollable list box. The "Estimation Settings:" section has a "Method:" dropdown menu set to "LS - Least Squares (NLS and ARMA)" and a "Sample:" text box containing "1952:1 1992:4". On the right side, there are three buttons: "OK" (with a green checkmark), "Cancel" (with a red X), and "Options" (with a yellow icon).



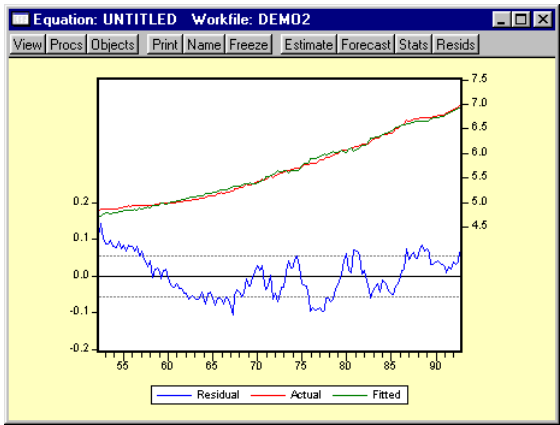
Here we list the name of the dependent variable, followed by the names of each of the regressors, separated by spaces. We use expressions involving the functions `log` and `dlog` to represent the log transformations of M1 and GDP, and the difference of the log transformation for PR. The built-in series name C stands for the constant in the regression.

The dialog is initially set to estimate the equation using the **LS - Least Squares** method for the **Sample** 1952:1 1996:4. You should change the **Sample** to 1952:1 1992:4 to estimate the equation for the subsample of observations.

Click **OK** to estimate the equation using least squares and to display the regression results:

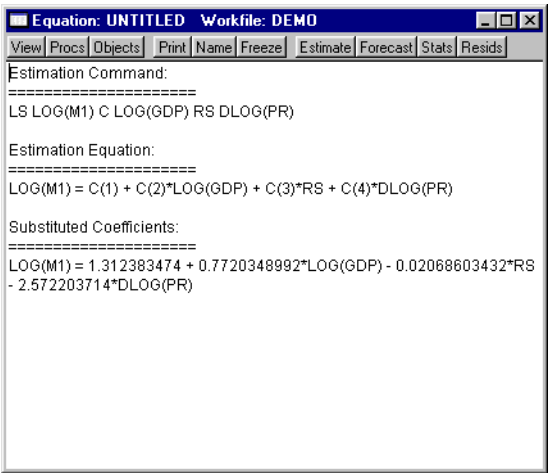
Dependent Variable: LOG(M1)				
Method: Least Squares				
Date: 10/19/97 Time: 22:43				
Sample(adjusted): 1952:2 1992:4				
Included observations: 163 after adjusting endpoints				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.312383	0.032199	40.75850	0.0000
LOG(GDP)	0.772035	0.006537	118.1092	0.0000
RS	-0.020686	0.002516	-8.221196	0.0000
DLOG(PR)	-2.572204	0.942556	-2.728967	0.0071
R-squared	0.993274	Mean dependent var		5.692279
Adjusted R-squared	0.993147	S.D. dependent var		0.670253
S.E. of regression	0.055485	Akaike info criterion		-2.921176
Sum squared resid	0.489494	Schwarz criterion		-2.845256
Log likelihood	242.0759	F-statistic		7826.904
Durbin-Watson stat	0.140967	Prob(F-statistic)		0.000000

Note that the equation is estimated from 1952:2 to 1992:4 since one observation is dropped from the beginning of the estimation sample to account for the `dlog` difference term. The estimated coefficients are statistically significant, with *t*-statistic values well in excess of 2. The overall regression fit, as measured by the  $R^2$  value, indicates a very tight fit. You can select **View/Actual, Fitted, Residual/Graph** in the equation toolbar to display a graph of the actual and fitted values for the dependent variable, along with the residuals:



### Specification and Hypothesis Tests

We can use the estimated equation to perform hypothesis tests on the coefficients of the model. For example, to test the hypothesis that the coefficient on the price term is equal to 2, we will perform a Wald test. First, determine the coefficient of interest by selecting **View/Representations** from the equation toolbar:



Note that the coefficients are assigned in the order that the variables appear in the specification so that the coefficient for the PR term is labeled C(4). To test the restriction on C(4) you should select **View/Coefficient Tests/Wald–Coefficient Restrictions...**, and enter the restriction  $c(4) = 2$ . EViews will report the results of the Wald test:

Wald Test:			
Equation: Untitled			
Null Hypothesis: C(4)=2			
F-statistic	23.53081	Probability	0.000003
Chi-square	23.53081	Probability	0.000001

The low probability values indicate that the null hypothesis that  $C(4) = 2$  is strongly rejected.

We should, however, be somewhat cautious of accepting this result without additional analysis. The low value of the Durbin-Watson statistic reported above is indicative of the presence of serial correlation in the residuals of the estimated equation. If uncorrected, serial correlation in the residuals will lead to incorrect estimates of the standard errors, and invalid statistical inference for the coefficients of the equation.

The Durbin-Watson statistic can be difficult to interpret. To perform a more general Breusch-Godfrey test for serial correlation in the residuals, select **View/Residual Tests/Serial Correlation LM Test...** from the equation toolbar, and specify an order of serial correlation to test against. Entering 1 yields a test against first-order serial correlation:

Breusch-Godfrey Serial Correlation LM Test:				
F-statistic	813.0060	Probability	0.000000	
Obs*R-squared	136.4770	Probability	0.000000	
Test Equation:				
Dependent Variable: RESID				
Method: Least Squares				
Date: 10/19/97 Time: 22:45				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.006355	0.013031	-0.487683	0.6265
LOG(GDP)	0.000997	0.002645	0.376929	0.7067
RS	-0.000567	0.001018	-0.556748	0.5785
DLOG(PR)	0.404143	0.381676	1.058864	0.2913
RESID(-1)	0.920306	0.032276	28.51326	0.0000
R-squared	0.837282	Mean dependent var	1.21E-15	
Adjusted R-squared	0.833163	S.D. dependent var	0.054969	
S.E. of regression	0.022452	Akaike info criterion	-4.724644	
Sum squared resid	0.079649	Schwarz criterion	-4.629744	
Log likelihood	390.0585	F-statistic	203.2515	
Durbin-Watson stat	1.770965	Prob(F-statistic)	0.000000	

The top part of the output presents the test statistics and associated probability values. The test regression used to carry out the test is reported below the statistics.

The statistic labeled “Obs\*R-squared” is the LM test statistic for the null hypothesis of no serial correlation. The (effectively) zero probability value strongly indicates the presence of serial correlation in the residuals.

## Modifying the Equation

The test results suggest that we need to modify our original specification to take account of the serial correlation.

One approach is to include lags of the independent variables. To add variables to the existing equation, click on the **Estimate** button in the equation toolbar and edit the specification to include lags for each of the original explanatory variables:

```
log(m1) c log(gdp) rs dlog(pr) log(m1(-1)) log(gdp(-1)) rs(-1)
      dlog(pr(-1))
```

Note that lags are specified by including a negative number, enclosed in parentheses, following the series name. Click on **OK** to estimate the new specification and to display the results:

Dependent Variable: LOG(M1)  
Method: Least Squares  
Date: 10/19/97 Time: 22:48  
Sample(adjusted): 1952:3 1992:4  
Included observations: 162 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.071297	0.028248	2.523949	0.0126
LOG(GDP)	0.320338	0.118186	2.710453	0.0075
RS	-0.005222	0.001469	-3.554801	0.0005
DLOG(PR)	0.038615	0.341619	0.113036	0.9101
LOG(M1(-1))	0.926640	0.020319	45.60375	0.0000
LOG(GDP(-1))	-0.257364	0.123264	-2.087910	0.0385
RS(-1)	0.002604	0.001574	1.654429	0.1001
DLOG(PR(-1))	-0.071650	0.347403	-0.206246	0.8369
R-squared	0.999604	Mean dependent var	5.697490	
Adjusted R-squared	0.999586	S.D. dependent var	0.669011	
S.E. of regression	0.013611	Akaike info criterion	-5.707729	
Sum squared resid	0.028531	Schwarz criterion	-5.555255	
Log likelihood	470.3261	F-statistic	55543.30	
Durbin-Watson stat	2.393764	Prob(F-statistic)	0.000000	

Note that EViews has automatically adjusted the estimation sample to accommodate the additional lagged variables. We will save this equation in the workfile for later use. Press the **Name** button in the toolbar and name the equation EQLAGS.

Another common method of accounting for serial correlation is to include autoregressive (AR) and/or moving average (MA) terms in the equation. To estimate the model with an AR(1) error specification, you should make a copy of the previous equation by clicking **Objects/Copy Object...** EViews will create a new untitled equation containing all of the information from the

previous equation. Press **Estimate** on the toolbar of the copy and modify the specification to read

```
log(m1) c log(gdp) rs dlog(pr) ar(1)
```

This specification removes the lagged terms, replacing them with an AR(1) specification. Click **OK**. EViews will report the estimation results, including the estimated first-order autoregressive coefficient of the error term:

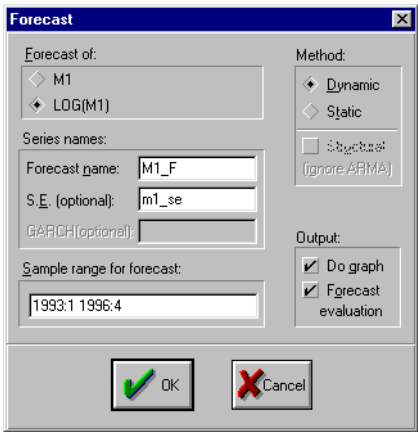
Dependent Variable: LOG(M1)				
Method: Least Squares				
Date: 10/19/97 Time: 22:52				
Sample(adjusted): 1952:3 1992:4				
Included observations: 162 after adjusting endpoints				
Convergence achieved after 14 iterations				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.050340	0.328390	3.198453	0.0017
LOG(GDP)	0.794929	0.049342	16.11057	0.0000
RS	-0.007395	0.001457	-5.075131	0.0000
DLOG(PR)	-0.008019	0.348689	-0.022998	0.9817
AR(1)	0.968100	0.018190	53.22283	0.0000
R-squared	0.999526	Mean dependent var	5.697490	
Adjusted R-squared	0.999514	S.D. dependent var	0.669011	
S.E. of regression	0.014751	Akaike info criterion	-5.564584	
Sum squared resid	0.034164	Schwarz criterion	-5.469288	
Log likelihood	455.7313	F-statistic	82748.93	
Durbin-Watson stat	2.164265	Prob(F-statistic)	0.000000	
Inverted AR Roots	.97			

The fit of the AR(1) model is roughly comparable to the lag model, but the somewhat higher values for both the Akaike and the Schwarz information criteria indicate that the previous lag model should be preferred. We will work with the lag model for the remainder of the demonstration.

## Forecasting from an Estimated Equation

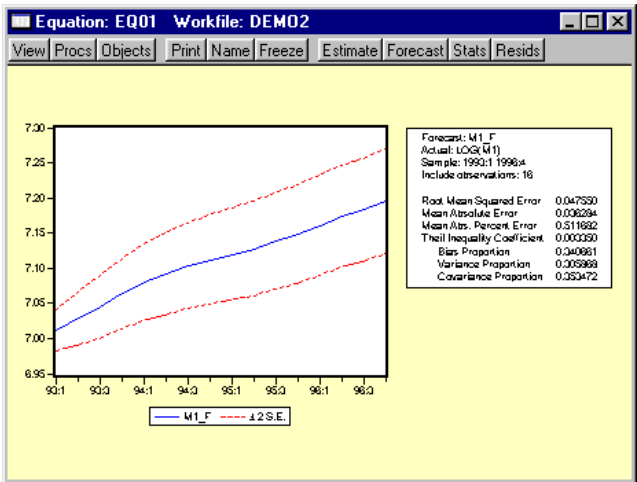
We have estimated the equations for a subset of our data, so that we may compare forecasts based upon this model with the actual data for the post-estimation sample 1993:1–1996:4.

Click on the **Forecast** button in the EQLAGS equation toolbar to open the forecast dialog:

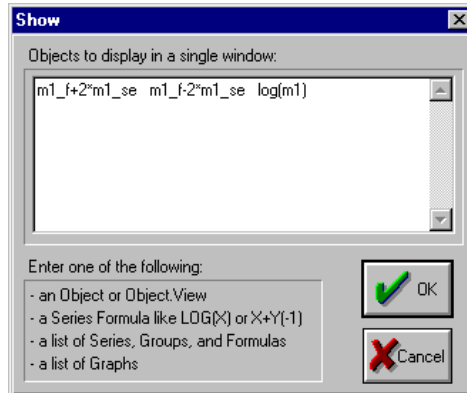


We set the forecast sample to 1993:1–1996:4 and provide names for both the forecasts and forecast standard errors so both will be saved as series in the workfile. The forecasted values will be saved in M1\_F and the forecast standard errors will be saved in M1\_SE.

Note also that we have elected to forecast the log of M1, not the level, and that we request both graphical and forecast evaluation output. The **Dynamic** option constructs the forecast for the sample period using only information available at the beginning of 1993:1. When you click **OK**, EViews displays both a graph of the forecasts, and statistics evaluating the quality of the fit to the actual data:

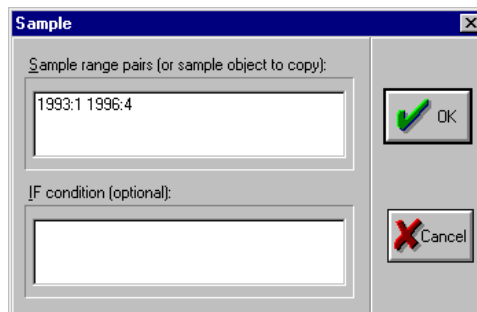


We can also plot the actual values of  $\log(M1)$  against the forecasted values and the (approximate) 95% confidence intervals for the forecasts. First, we will create a new group containing these values by **Quick/Show...** and filling out the dialog as follows:

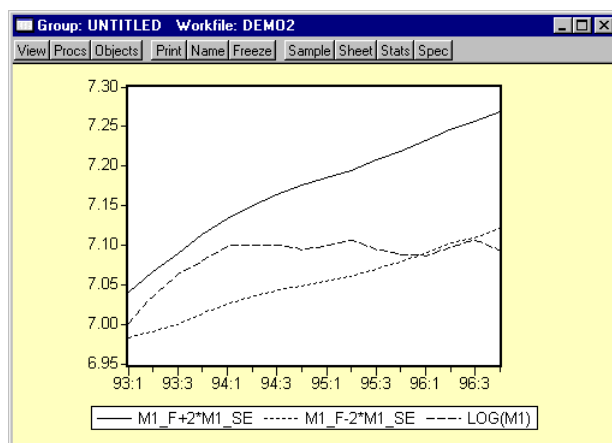


There are three expressions in the dialog. The first two represent the upper and lower bounds of the (approximate) 95% forecast interval as computed by evaluating the values of the point forecasts plus and minus two times the standard errors. The last expression represents the actual values of the dependent variable.

When you click **OK**, EViews opens an untitled group window containing a spreadsheet view of the data. Before plotting the data, we will change the sample of observations so that we only plot data for the forecast sample. Select **Quick/Sample...** or click on the **Sample** button in the group toolbar, and change the sample to include only the forecast period:

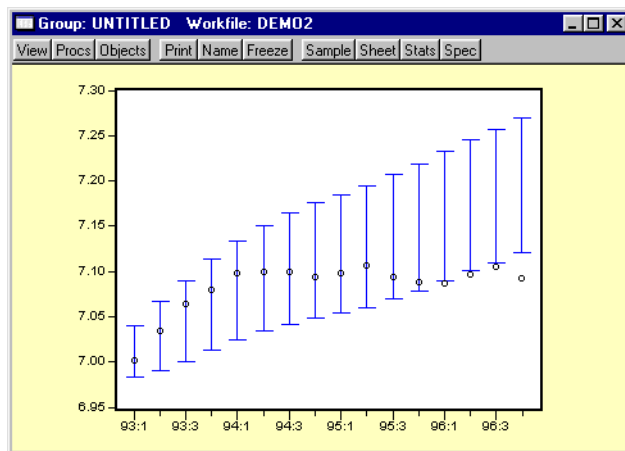


To plot the data for the forecast period, select **View/Graph/Line** from the group window:



The actual values of  $\log(M1)$  are within the forecast interval for most of the forecast period, but fall below the lower bound of the 95% confidence interval beginning in 1996:1.

For an alternate view of these data, you can select **View/Graph/High-Low(-Close)**, which displays the graph as follows:

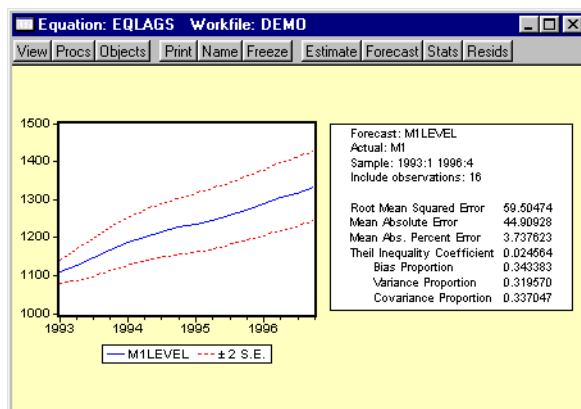


This high-low chart clearly shows that the forecasts of  $\log(M1)$  over-predict the actual values in the last four quarters of the forecast period.

We may also choose to examine forecasts of the level of  $M1$ . Click on the **Forecast** button in the EQLAGS equation toolbar to open the forecast dialog, and select  $M1$  under the **Forecast of** option. Enter a new name to hold the forecasts, say  $M1LEVEL$ , and click **OK**. EVIEWS will present



a graph of the forecast of the level of M1, along with the asymmetric confidence intervals for this forecast:



## Additional Issues

It is worth noting that the example analysis above should be used for illustrative purposes only since there are a number of problems with the specification.

For one, there is quite a bit of serial correlation remaining in the EQLAGS specification. A test of serial correlation in the new equation (by selecting **View/Residual Tests/Serial Correlation LM Test...**, and entering 1 for the number of lags) rejects the null hypothesis of no serial correlation in the reformulated equation:

### Breusch-Godfrey Serial Correlation LM Test:

F-statistic	7.880369	Probability	0.005648
Obs*R-squared	7.935212	Probability	0.004848

Furthermore, there is evidence of autoregressive conditional heteroskedasticity (ARCH) in the residuals. Select **View/Residual Tests/ARCH LM Test...** and accept the default of 1. The ARCH test results strongly suggest the presence of ARCH in the residuals:

### ARCH Test:

F-statistic	11.21965	Probability	0.001011
Obs*R-squared	10.61196	Probability	0.001124

In addition to serial correlation and ARCH, there is an even more fundamental problem with the above specification since, as the graphs attest,  $\log(M1)$  exhibits a pronounced upward trend. We

can, and should, perform tests for a unit root in this series. The presence of a unit root will indicate the need for further analysis.

Display the series window by clicking on **Window** and selecting the LOG(M1) series window from the menu. If the series window is closed, you may open a new window by selecting **Quick/Show...**, entering `log(m1)`, and clicking OK.

To perform an Augmented Dickey-Fuller (ADF) test for nonstationarity of this series, select **View/Unit Root Test...** and click on OK to accept the default options. EViews will perform an ADF test and display the test results:

ADF Test Statistic	0.665471	1% Critical Value*	-3.4688
		5% Critical Value	-2.8780
		10% Critical Value	-2.5755

\*MacKinnon critical values for rejection of hypothesis of a unit root.

The ADF test statistic value is greater than the critical values so that we cannot reject the null hypothesis of a unit root. The presence of a unit root suggests that we need to adopt more sophisticated statistical models. These techniques are discussed in Chapters 13 and 20 of the *User's Guide* which deal with time series and vector autoregression and vector error correction specifications.

## Chapter 3. EViews Basics

---

Managing the variety of tasks associated with your work can be a complex and time-consuming process. Fortunately, EViews' innovative design takes much of the effort out of organizing your work, allowing you to concentrate on the substance of your project.

At the heart of the EViews design is the concept of an object. In brief, objects are collections of related information and operations that are bundled together into an easy-to-use unit. Virtually all of your work in EViews will involve using and manipulating various objects.

EViews holds all of its objects in object containers. You can think of object containers as filing cabinets or organizers for the various objects with which you are working. The most important object container in EViews is the workfile.

The remainder of this chapter describes basic techniques for working with objects and workfiles. While you may at first find the idea of objects to be a bit foreign, the basic concepts are easy to master and will form the foundation for your work in EViews. But don't feel that you have to understand all of the concepts the first time through. If you wish, you can begin working with EViews immediately, developing an intuitive understanding of objects and workfiles as you go.

Subsequent chapters will provide a more detailed description of working with the various types of objects and other types of object containers.

### Workfile Basics

All EViews objects must be held in an object container. Most of your work in EViews will involve objects that are contained in a workfile, so your first step in any project will be to create a new workfile or to load an existing workfile into memory.

Workfiles have two primary characteristics. First, they are held in RAM for quick access to the objects in the workfile. Second, workfiles are characterized by a frequency and a range.

Data are often sampled at equally spaced intervals, or frequencies, over calendar time. When you set a workfile frequency you tell EViews about the intervals between observations in your data. EViews has dated workfiles which handle annual, semi-annual, quarterly, monthly, weekly, and daily (5- or 7-day) data. For these types of workfiles, EViews will use all available calendar information in organizing and managing your data. For example, for weekly and daily data, EViews knows that some years contain

days in each of 53 weeks, and that some years have 366 days, and will adjust the number of observations in a year accordingly.

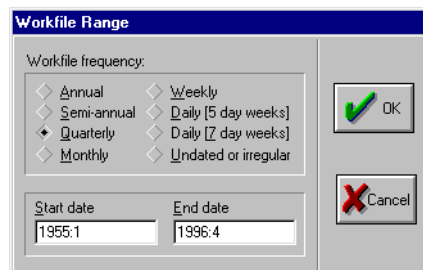
Undated or irregular workfiles are those in which no dates are associated with the data—observations are simply numbered consecutively. Undated data are typically used for cross-section data, but may also be used in any situation where data are sampled irregularly; for example, financial data with frequent and irregular breaks for non-trading days.

The workfile range is a pair of dates or observation numbers describing the first and last observation to be held in the workfile.

## Creating a Workfile

Your first step in EViews will usually be to create a workfile. One way to create a workfile is to click **File/New/Workfile...** and then to provide the necessary dialog information.

Select the appropriate frequency and enter the information for the workfile range. The **Start date** is the earliest date or observation you plan to use in the project and the **End date** is the latest date or observation. Don't worry if you don't know the exact start and end date; if you later find that your workfile isn't big enough, you can expand the workfile range.



The rules for describing dates are quite simple:

**Annual:** specify the year. Years before and after the 20th century must be identified with full-year identifiers (e.g. 1776, 2020, 9789 or 50234). Years in the 20th century may be identified using either 2 or 4-digit identifiers (e.g. 97 or 1997). Note that because 2-digit identifiers are assumed to be in the 20th century, EViews cannot handle dates prior to A.D. 100.

**Quarterly:** the year, followed by a colon or period, and the quarter number. Examples: 1992:1, 65:4, 2002:3.

**Monthly:** the year, followed by a colon or period, and the month number. Examples: 1956:1, 1990:11.

**Weekly and daily:** by default, you should specify these dates as month number, followed by a colon, followed by the day number, followed by a colon, followed by the year. Using the **Options/Dates-Frequency...** menu item, you can reverse the order of the day and month by switching to European notation.

For example, entering 8:10:97 indicates that you want your workfile to begin with August 10, 1997. If you set your default to European notation, the desired starting date is October 8, 1997.

With weekly data, the day of the week associated with the starting date determines the

beginning of the week. In the examples above, the first observations would be the week running from Sunday, August 10 through Saturday, August 16, 1997, or the week running from Wednesday, October 8, through Tuesday, October 14, 1997.

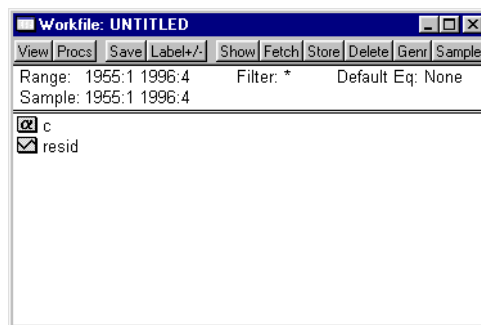
Alternatively, for quarterly, monthly, weekly, and daily data, you can enter just the year, and EViews will automatically specify the first and last observations for you.

In Appendix C, “Date Formats”, we discuss the specification of dates in EViews in greater detail.

After you have finished supplying the information about the type of workfile and clicked **OK**, you will see the workfile window:

Here we have specified a workfile which will contain quarterly data from the first quarter of 1955 through the end of 1996. Since we have not yet saved the workfile, it is UNTITLED.

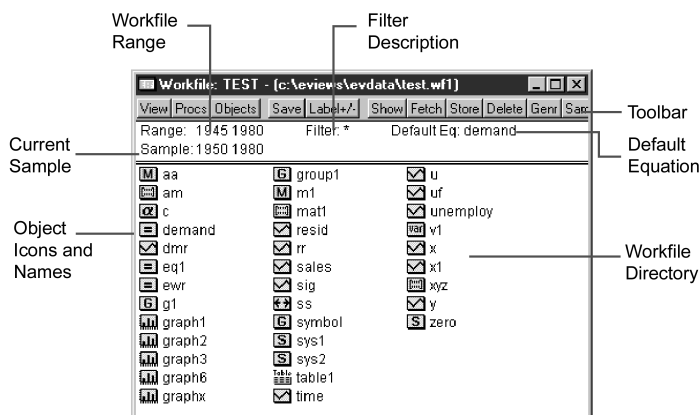
Note that there are two icons in this newly created workfile. These icons represent the objects that are contained in *every* workfile: a vector of coefficients, C, and a series of residuals, RESID. The little icon to the left identifies the type of object, an  $\alpha$  for a coefficient vector and a tiny time series plot for a series.



Workfiles may also be created directly from EViews databases. See Chapter 6 for further details.

## The Workfile Window

After you have created a workfile and a number of objects, the workfile window will look something like this:



In the titlebar of the workfile window you will see the “**Workfile**” designation followed by the workfile name. If the workfile has not been saved, it will be designated “UNTITLED”. If the workfile has been saved to disk, you will see the name and the full disk path.

Just below the titlebar is a toolbar made up of a number of buttons. These buttons provide you with easy access to a number of useful workfile operations.

Below the toolbar are two lines of status information where EViews displays the range of the workfile, the current *sample* of the workfile (the range of observations that are to be used in calculations and statistical operations), the display filter (rule used in choosing a subset of objects to display in the workfile window), and the *default equation* (the last equation estimated or operated on). You may change the range, sample, and filter by double clicking on these labels and entering the relevant information in the dialog boxes. Double clicking on the equation label opens the equation.

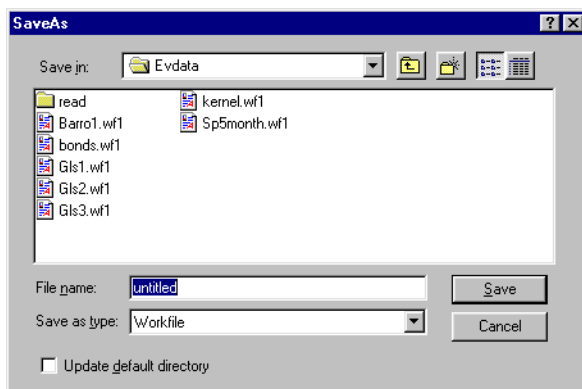
Lastly, you will see the workfile directory. In normal display mode, all named objects are listed in the directory by name and icon. The different types of objects and their icons are described in detail in “Object Types” on page 42.

It is worth remembering that the workfile window is a specific example of an object window. Object windows are discussed in “The Object Window” on page 45.

## Saving Workfiles

You will want to name and save your workfile for future use. Push the **Save** button on the workfile toolbar to save a copy of the workfile on disk. You can also save the file using the **File/SaveAs...** or **File/Save...** choices from the main menu. A standard Windows file dialog box will open:

You can specify the target directory in the upper file menu labeled **Save in:**. You can navigate between directories in the standard Windows fashion—click once on the down arrow to access a directory tree; double clicking on a directory name in the display area gives you a list of all the files and subdirectories in that directory. Once you have worked your way to the right directory, type the name you want to give the workfile in the **File name:** box and push the **Save** button. Your workfile will be saved with the name you choose and the extension .WF1.



Alternatively, you could just type the full Windows path information and name, in the **File name:** box.

Once the workfile is named and saved, you can save subsequent updates or changes using **File/Save...** from the main menu or the **Save** button on the toolbar. EViews will use the existing workfile name, and if the workfile has changed, will ask you whether you want to update the version on disk. As with other Windows software, **File/Save As...** can be used to save the file with a new name.

Note that workfiles saved in version 3 can, in general, be read by previous versions of EViews. Objects that are new to version 3 will, however, be removed from the workfile. We recommend that you take great caution when saving over your workfile using older versions of EViews.

## Loading Workfiles

You can use **File/Open/Workfile...** to bring back a previously saved workfile. You will typically save your workfile containing all of your data and results at the end of the day, and later use **File/Open/Workfile...** to pick up where you left off.

When you select **File/Open/Workfile...** you will see a standard Windows file dialog. Simply navigate to the appropriate directory and double click on the name of the workfile to load it into RAM. The workfile window will open and all of the objects in the workfile will immediately be available.

For your convenience, EViews keeps a record of the ten most recently used workfiles and programs at the bottom of the File menu. Select an entry and it will be opened in EViews.

Version 3 of EViews can read workfiles from all previous versions of EViews.

## Save and Load Options

There are optional settings in the **File/Open...** and **File/Save As...** dialogs which provide you with additional control over the procedures which use files saved on disk.

### Set Default Directory

All EViews file dialogs begin with a display of the contents of the *default directory*. You can always identify the default directory from the listing on the EViews status line. The default directory is set initially to be the directory containing the EViews program, but it can be changed at any time.

You can change the default directory by using the **File/Open...** or the **File/SaveAs...** menu items, navigating to the new directory, and checking the **Update Default Directory** box in the dialog. If you then open or save a workfile, the default directory will change to the one you have selected.

An alternative method for changing the default EViews directory is to use the `cd` command. Simply enter “`cd`” followed by the directory name in the command window (see the *Command and Programming Reference* for details).

### Using MicroTSP Files

You can read or write your workfile in a format that is compatible with MicroTSP. The **Files of Type:** and **Save as Type:** drop boxes allow you to handle DOS and Macintosh MicroTSP files. Simply click on the drop down box and select either **Old Dos Workfile** or **Old Mac Workfile**, as appropriate. You should be aware, however, that if you choose to save a workfile in MicroTSP format, only basic series data will be saved—the remainder of the workfile contents will be discarded.

### Resizing Workfiles

You may decide to add data or make forecasts for observations beyond the ending date or before the starting date of your workfile. Alternatively, you may wish to remove extra observations from the start or end of the workfile.

To change the size of your workfile, select **Procs/Change workfile Range...** and enter the beginning and ending observation of the workfile in the dialog. If you enter dates that encompass the original workfile range, EViews will expand the workfile without additional comment. If you enter a workfile range that does not encompass the original workfile range, EViews will warn you that data will be lost, and ask you to confirm the operation.

### Sorting Workfiles

Basic data in workfiles are held in objects called series. If you click on **Procs/Sort Series...** in the workfile toolbar, you can sort all of the series in the workfile on the basis of the values of one or more of the series. A dialog box will open where you can provide the details about the sort.

If you list two or more series, EViews uses the values of the second series to resolve ties from the first series, and values of the third series to resolve ties from the second, and so forth. If you wish to sort in descending order, select the appropriate option in the dialog.

Note that if you are using a dated workfile, sorting the workfile will generally break the link between an observation and the corresponding date.

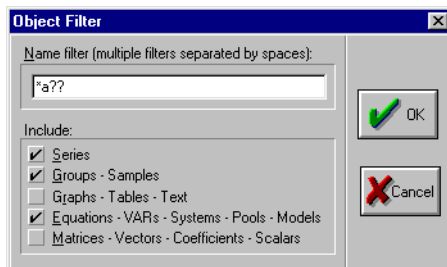
### Changing the Workfile Display

#### Display Filter

When working with workfiles containing a large number of objects, it can become difficult to locate specific objects in the workfile window. You can solve this problem by using the workfile display filter to instruct EViews to display only a subset of objects in the workfile window. This subset can be defined on the basis of object name as well as object type.



Select **View/Display Filter...** or double click on the Filter description in the workfile window. The following dialog box will appear:



There are two parts to this dialog. In the edit field (blank space) of this dialog, you may place one or several name descriptions that include the standard wildcard characters: “\*” (match any number of characters) and “?” (match any single character). Below the edit field are a series of check boxes corresponding to various types of EViews objects. EViews will display only objects of the specified types whose names match those in the edit field list.

The default string is “\*”, which will display all objects of the specified types. However, if you enter the string

x\*

only objects with names beginning with X will be displayed in the workfile window. Entering

x?y

displays all objects that begin with the letter X, followed by any single character and then ending with the letter Y. If you enter:

x\* y\* \*z

all objects with names beginning with X or Y and all objects with names ending in Z will be displayed. Similarly, the more complicated expression:

??y\* \*z\*

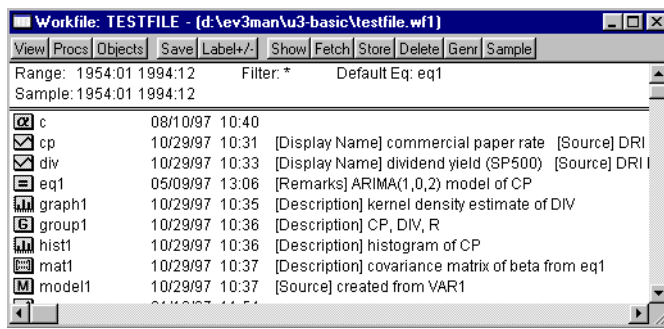
tells EViews to display all objects that begin with any two characters followed by a Y and any or no characters, and all objects that contain the letter Z. Wildcards may also be used in more general settings—a complete description of the use of wildcards in EViews is provided in Appendix D.

When you specify a display filter, the Filter description in the workfile window changes to reflect your request. EViews always displays the current string used in matching names. Additionally, if

you have chosen to display a subset of EViews object types, a “–” will be displayed in the Filter description at the top of the workfile window.

### Display Comments

You can change the default workfile display to show additional information about your objects. If you select **View/Display Comments (Label + –)**, EViews will toggle between the standard workfile display format, and a display which provides additional information about the date the object was created or updated, as well as the label information that you may have attached to the object.



### Display Letter Format

You can choose **View/Name Display...** in the workfile toolbar to specify whether EViews should use upper or lower case letters when it displays the workfile directory. The default is lower case.

## Object Basics

Information in EViews is stored in *objects*. Each object consists of a collection of information related to a particular area of analysis. For example, a *series object* is a collection of information related to a set of observations on a particular variable. An *equation object* is a collection of information related to the relationship between a collection of variables.

Note that an object need not contain only one type of information. For example, an estimated equation object contains not only the coefficients obtained from estimation of the equation, but also a description of the specification, the variance-covariance matrix of the coefficient estimates, and a variety of statistics associated with the estimates.

Associated with each type of object is a set of views and procedures which can be used with the information contained in the object. This association of views and procedures with the type of data contained in the object is what we term the *object oriented* design of EViews.

The object oriented design simplifies your work in EViews by organizing information as you work. For example, since an equation object contains all of the information relevant to an esti-

mated relationship, you can move freely between a variety of equation specifications simply by working with different equation objects. You can examine results, perform hypothesis and specification tests, or generate forecasts at any time. Managing your work is simplified since only a single object is used to work with an entire collection of data and results.

This brief discussion provides only the barest of introduction to the use of objects. The remainder of this section will provide a more general description of EViews objects. Subsequent chapters will discuss series, equations, and other object types in considerable detail.

## Object Data

Each object contains various types of information. For example, series, matrix, vector, and scalar objects, all contain mostly numeric information. In contrast, equations and systems contain complete information about the specification of the equation or system, and the estimation results, as well as references to the underlying data used to construct the estimates. Graphs and tables contain numeric, text, and formatting information.

Since objects contain various kinds of data, you will want to work with different objects in different ways. For example, you might wish to compute summary statistics for the observations in a series, or you may want to perform forecasts based upon the results of an equation. EViews understands these differences and provides you with custom tools, called views and procedures, for working with an object's data.

## Object Views

There is more than one way to examine the data in an object. Views are tabular and graphical windows that provide various ways of looking at the data in an object.

For example, a series object has a spreadsheet view, which shows the raw data, a line graph view, a bar graph view, a histogram-and-statistics view, and a correlogram view. Other views of a series include distributional plots, QQ-plots, and kernel density plots. Series views also allow you to compute simple hypothesis tests and statistics for various subgroups of your sample.

An equation object has a representation view showing the equation specification, an output view containing estimation results, an actual-fitted-residual view containing plots of fitted values and residuals, a covariance view containing the estimated coefficient covariance matrix, and various views for specification and parameter tests.

Views of an object are displayed in the object's window. Only one window can be opened for each object and each window displays only a single view of the object at a time. You can change views of an object using the View menu located in the object window's toolbar or the EViews main menu. We will describe this process in greater detail below.

Perhaps the most important thing to remember about views is that changing the view of an object does not in any way change the data in the object. The underlying data in the object remains the same—only the display of the data changes.

## Object Procedures



















Most EViews objects also have *procedures*, or *procs*. Like views, procedures often display tables or graphs in the object's window. Unlike views, however, procedures alter data, either in the object itself or in another object.

Many procedures create new objects. For example, a series object contains procedures for smoothing or seasonally adjusting time series data and creating a new series containing the smoothed or adjusted data. Equation objects contain procedures for generating new series containing the residuals, fitted values, or forecasts from the estimated equation.

You select procedures from the **Procs** menu on the object's toolbar or from the EViews menu.

## Object Types

The most fundamental objects in EViews are series and equation objects. There are, however, a number of different types of objects, each of which serves a unique function. Most objects are represented by a unique icon which is displayed in the object container window:

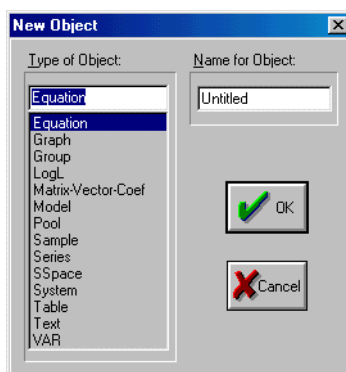
	Coefficient Vector		Scalar
	Equation		Series
	Graph		State Space
	Group		System
	LogL		SYM (Symmetric Matrix)
	Matrix		Table
	Model		Text
	Pool (Time Series / Cross-Section)		VAR (Vector Autoregression)
	Sample		Vector/Row Vector

Despite the fact that they are also objects, object containers do not have icons since they cannot be placed in other object containers—thus, workfiles and databases do not have icons since they cannot be placed in other workfiles or databases.

## Creating, Selecting, and Opening Objects

### Creating Objects

To create an object, you must first make certain that you have an open workfile container and that the workfile window is active. Next, select **Objects/New Object...** from the main menu. Until you have created or loaded a workfile, this selection is unavailable. After you click on the menu entry, you will see the following dialog box:



You can click on the type of object you want, optionally provide a name and then click on **OK**. For some object types, another dialog box will open prompting you to describe your object in more detail. For most objects, however, the object window will open immediately.

For example, if you select **Equation**, you will see a dialog box prompting you for additional information. Alternatively, if you click on **Series** and then select **OK**, you will see an object window (series window) displaying the spreadsheet view of an UNTITLED series:

 A screenshot of the 'Series: UNTITLED' window. The title bar says 'Series: UNTITLED Workfile: UNTITLED'. Below the title bar is a menu bar with 'View', 'Procs', 'Objects', 'Print', 'Name', 'Freeze', 'Edit+/-', 'Smpl+/-', 'Label+/-', 'InsDel', 'Title', and 'Sam'. Below the menu bar is a spreadsheet grid. The grid has columns labeled 1, 2, 3, and 4. The first column is labeled '1955:1' through '1964:1'. The first four rows of data are highlighted. The data in the first four rows is: 1955:1 (NA, NA, NA, NA), 1956:1 (NA, NA, NA, NA), 1957:1 (NA, NA, NA, NA), 1958:1 (NA, NA, NA, NA). The status bar at the bottom says 'Last updated: 08/05/97 - 23:00'.
 

	1	2	3	4
1955:1	NA	NA	NA	NA
1956:1	NA	NA	NA	NA
1957:1	NA	NA	NA	NA
1958:1	NA	NA	NA	NA
1959:1	NA	NA	NA	NA
1960:1	NA	NA	NA	NA
1961:1	NA	NA	NA	NA
1962:1	NA	NA	NA	NA
1963:1	NA	NA	NA	NA
1964:1				

We will discuss object windows in greater detail below.

Objects can also be created by applying procedures to other objects or by freezing an object view (see “Freezing Objects” on page 49, below).

### Selecting Objects

Creating a new object will not always be necessary. Instead, you may want to work with an existing object. One of the fundamental operations in EViews is *selecting* one or more objects from the workfile directory.

The easiest way to select objects is to point-and-click, using the standard Windows conventions for selecting contiguous or multiple items if necessary (see Chapter 1, “Selecting and Opening Items” on page 8). Keep in mind that if you are selecting a large number of items, you may find it useful to use the display filter before beginning to select items.

In addition, the **View** button in the workfile toolbar provides convenient selection shortcuts:

- **Deselect All** eliminates any existing selections.
- **Select All** selects all of the objects in the workfile with the exception of the C coefficient vector and the RESID series.

### Opening Objects

Once you have selected your object or objects, you will want to open your selection, or create a new object containing the selected objects. You can do so by double clicking anywhere in the highlighted area.

If you double click on a single selected object, you open an object window.

If you select multiple graphs or series and double click, a pop-up menu appears giving you the option of creating and opening new objects (group, equation, VAR, graph) or displaying each of the selected objects in its own window.

Note that if you select multiple graphs and double click or select **View/Open as One Window**, all of the graphs are merged into a single graph that is displayed in a window.

Other multiple item selections are not valid, and will either issue an error or will simply not respond when you double click.

When you open an object, EViews will display the default view. In general, the default view of an object is the view that was displayed the last time the object was opened (if an object has never been opened, EViews will use a default view). The exception to this general rule is for those views that require significant computational time. In this latter case, the default view will revert to the new object default.

### Showing Objects

An alternative method of selecting and opening objects is to “show” the item. Click on the **Show** button on the toolbar, or select **Quick/Show...** from the menu and type in the object name or names.

Show works exactly as if you first selected the object or objects, and then opened your selection. If you enter a single object name in the dialog box, EViews will open the object as if you double clicked on the object name. If you enter multiple names, EViews will always open a single window to display results, creating a new object if necessary.

Show can also be used to display functions of series, also known as auto-series. All of the rules for auto-series outlined in Chapter 5, “Working with Auto-series” on page 92, will apply.

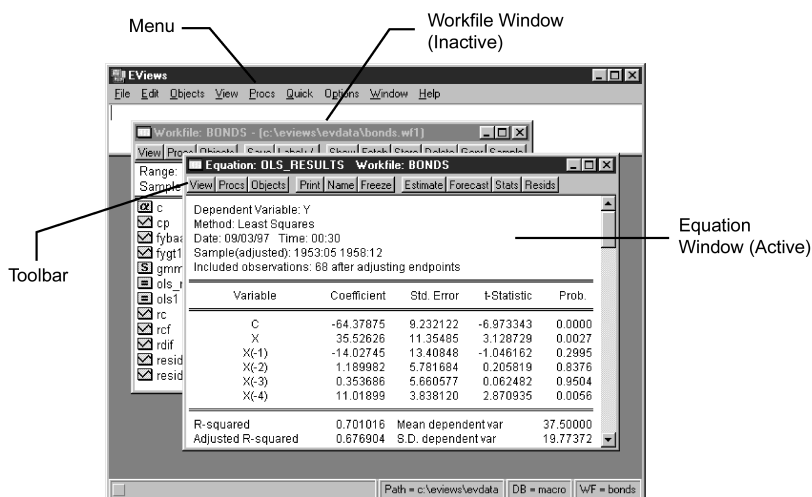
## The Object Window

We have been using the term object window somewhat loosely in the previous discussion of the process of creating and opening objects. Object windows are the windows that are displayed when you open an object or object container. An object’s window will contain either a view of the object, or the results of an object procedure.

One of the more important features of EViews is that you can display object windows for a number of items at the same time. Managing these object windows is similar to the task of managing pieces of paper on your desk.

## Components of the Object Window

Let’s look again at a typical object window:



Here, we see the equation window for OLS\_RESULTS. First, notice that this is a standard window which can be closed, resized, minimized, maximized, and scrolled both vertically and horizontally. As in other Windows applications, you can make an object window active by clicking once on the title bar, or anywhere in its window. Making an object window active is equivalent to saying that you want to work with that object.

Second, note that the title bar of the object window identifies the object type, name, and object container (in this case, the BONDS workfile). If the object is itself an object container, the container information is replaced by directory information.

Lastly, at the top of the window there is a toolbar containing a number of buttons that provide easy access to frequently used menu items. These toolbars will vary across objects—the series object will have a different toolbar from an equation or a group or a VAR object.

There are, however, several buttons that are found on all object toolbars:

- The **View** button lets you change the view that is displayed in the object window. The available choices will differ, depending upon the object type.
- The **Procs** button provides access to a menu of procedures that are available for the object.
- The **Objects** button lets you manage your objects. You can store the object on disk, name, delete, copy, or print the object.
- The **Print** button lets you print the current view of the object (the window contents).
- The **Name** button allows you to name or rename the object.
- The **Freeze** button creates a new object graph, table, or text object out of the current view.

The other buttons on the series toolbar are specific to a series object and are described in Chapter 7.

### Menus and the Object Toolbar

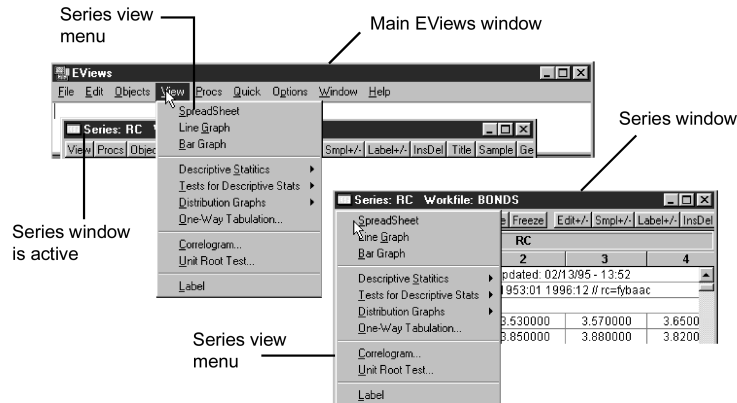
As we have seen, the toolbar provides a shortcut to frequently accessed menu commands. There are a couple of subtle, but important, points associated with this relationship that deserve special emphasis:

- Since the toolbar simply provides a shortcut to menu items, you can always find the toolbar commands in the menus.
- This fact turns out to be quite useful if your window is not large enough to display all of the buttons on the toolbar. You can either enlarge the window so that all of the buttons are displayed, or you can access the command directly from the menu.
- The toolbar and menu *both* change with the object type. In particular, the View, and Procs menu contents will always change to reflect the type of object (series, equation, group, etc.) that is active.

The toolbars and menus themselves vary in how much they differ across objects. For example, the View and Procs drop-down menus differ for every object type. When the active window is a series, the menus provide access to series views and series procedures. Alternatively, when the



active window is a group, clicking on **View** or **Procs** provides access to the different set of items associated with group objects.



The figure above illustrates the relationship between the View toolbar button and the View menu when the series is the active window. In the left side of the illustration, we see a portion of the *EViews* window, as it appears, after you click on View in the main menu (note that the RC series window is the active window). On the right, we see a depiction of the *series* window as it appears after you click on the **View** button in the series toolbar. Since the two operations are identical, the two drop-down menus are identical.

In contrast to the View and Procs Menus, the Objects menu does not, in general, vary across objects. An exception occurs, however, when an object container window (a workfile or data-base window) is active. In this case, clicking on **Objects** in the toolbar, or selecting **Objects** from the menu provides access to menu items for manipulating the objects in the container.

## Working with Objects

### Naming Objects

Objects may be named or unnamed. When you give an object a name, the name will appear in the directory of the workfile, and the object will be saved as part of the workfile when the workfile is saved.

You must name an object if you want to keep its results. If you do not name an object, it will be called “UNTITLED”. Unnamed objects are not saved with the workfile so that they are deleted when the workfile is closed and removed from memory.

To name or rename an object, first open the object window by double-clicking on its icon, or by clicking on **Show** on the workfile toolbar, and entering the object name. Next, click on the **Name** button on the object window, and enter the name (up to 16 characters), and optionally, a display

name to be used when referring to the object in tables and graphs. If no display name is provided, EViews will use the object name.

You can also rename an object from the workfile window by selecting **Objects/Rename Selected...** and then specifying the new object name. This method saves you from first having to open the object.

The following names are reserved and should not be used as object names: ABS, ACOS, AR, ASIN, C, CON, CNORM, COEF, COS, D, DLOG, DNORM, ELSE, ENDIF, EXP, LOG, LOGIT, LPT1, LPT2, MA, NA, NRND, PDL, RESID, RND, SAR, SIN, SMA, SQR, and THEN.

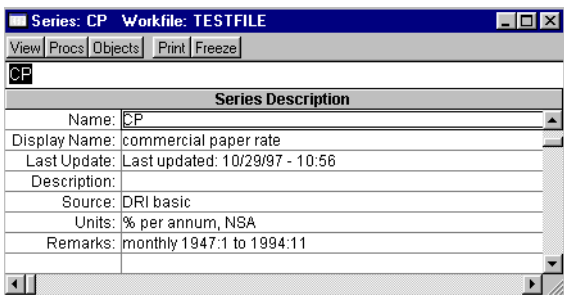
EViews accepts both capital letters and lower case letters in the names you give to your series and other objects, but it treats names as all capital letters, even if entered in lower case. Its messages to you will follow normal capitalization rules. For example, 'SALES', 'sales', and 'sAles' are all the same object in EViews. For the sake of uniformity, we have written all examples of input using names in lower case, but you should feel free to use capital letters instead.

Despite the fact that names are capitalized internally, when you enter text information in an object, such as a plot legend, or label information, your capitalization will be fully preserved.

By default, EViews allows only one untitled object of a given type (one series, one equation, etc.). If you create a new untitled object of an existing type, you will be prompted to name the original object, and if you do not provide one, EViews will replace the original untitled object with the new object. The original object will not be saved. If you prefer, you can instruct EViews to retain all untitled objects during a session (see Appendix B, "Window and Font Options" on page 579) but you must still name the ones you want to save with the workfile.

### Labeling Objects

In addition to the display name described above, EViews objects have label fields where you can provide extended annotation and commentary. To view these fields, select **View/Label** from the object window:



This is the label view of an unmodified object. Every time you modify the object, EViews by default automatically records the modification in a History field that will be appended at the bottom of the label view.

You can edit any of the fields, except the **Last Update** field. Simply click in the field cell that you want to edit. All fields, except the **Remarks** and **History** fields, contain only one

line. The **Remarks** and **History** fields can contain multiple lines. Press ENTER to add a new line to these two fields.

These annotated fields are most useful when you want to search an object stored in an EViews database. Any text that are in the fields are searchable in EViews databases; see Chapter 6, “Querying the Database”, page 120, for further discussion.

### Copying Objects

There are two distinct methods of duplicating the information in an object: copying and freezing.

If you select **Object/Copy** from the menu, EViews creates a new untitled object containing an exact copy of the original object. By exact copy, we mean that the new object duplicates all the features of the original (except for the name). It contains all of the views and procedures of the original object and can be used in future analyses just like the original object.

You can also copy an object from the workfile window. Simply highlight the object and click on **Object/Copy Selected...**, or click on **Object/Copy Selected...** and specify the destination name for the object.

We mention here that **Copy** is a very general and powerful operation with many additional features and uses. For example, you can copy objects across both workfiles and databases using wildcards and patterns. See Chapter 6, “Copy and Rename with Wildcard Characters” on page 110, for details on these additional features.

### Copy-and-Pasting Objects

The standard EViews copy command makes a copy of the object in the same workfile. When two workfiles are in memory at the same time, you may copy objects between them using Copy-and-Paste.

Highlight the objects you wish to copy in the source workfile. Then select **Edit/Copy** from the main menu.

Select the destination workfile by clicking on its title bar. Then select **Edit/Paste** from the main menu. EViews will place named copies of all of the highlighted objects in the destination workfile, prompting you to replace existing objects with the same name.

If the source and destination workfiles are of different frequency, frequency conversion (if possible) is applied to series objects before placing them in the destination workfile. See Chapter 6, “Frequency Conversion” on page 111, for the exact rules by which frequencies are converted.

### Freezing Objects

The second method of copying information from an object is to *freeze* a view of the object. If you click **Object/Freeze Output** or press the **Freeze** button on the object’s toolbar, a table or graph object is created that duplicates the current *view* of the original object.

Before you press **Freeze**, you are looking at a view of an object in the object window. Freezing the view makes a copy of the view and turns it into an independent object that will remain even

if you delete the original object. A frozen view does not necessarily show what is currently in the original object, but rather shows a snapshot of the object at the moment you pushed the button. For example, if you freeze a spreadsheet view of a series, you will see a view of a new *table object*; if you freeze a graphical view of a series, you will see a view of a new *graph object*.

The primary feature of freezing an object is that the tables and graphs created by freeze may be edited for presentations or reports. Frozen views do not change when the workfile sample or data change.

### Deleting Objects

To delete an object or objects from your workfile, select the object or objects in the workfile directory. When you have selected everything you want to delete, click **Delete** or **Objects/Delete Selected** on the workfile toolbar. EViews will prompt you to make certain that you wish to delete the objects.

### Printing Objects

Choosing **View/Print Selected** from the workfile window prints the default view for all of the selected objects.

To print the currently displayed view of an object, push the **Print** button on the object window toolbar. You can also choose **File/Print** or **Objects/Print** on the main EViews menu bar.

You may print the default view of more than one object at a time by selecting the objects in the workfile window and choosing **View/Print Selected** from the workfile toolbar.

The print commands normally send a view or procedure output to the current Windows printer. You may specify instead that the output should be saved in the workfile as a table or graph, or spooled to an ASCII text file on disk. Details are provided in Chapter 10, “Graphs, Tables, and Text Objects” and Appendix B, “Global Options”.

### Storing Objects

EViews provides three ways to save your data on disk. You have already seen how to save entire workfiles, where all of the objects in the workfile are saved together in a single file with the .WF1 extension. You may also store individual objects in their own *data bank* files. They may then be fetched into other workfiles.

We will defer a full discussion of storing objects to data banks and databases until Chapter 6. For now, note that when you are working with an object, you can place it in a data bank or database file by clicking on the **Objects/Store to DB...** button on the object's toolbar or menu. EViews will prompt you for additional information.

You can store several objects, by selecting them in the workfile window and then pressing the **Objects/Store selected to DB...** button on the workfile toolbar or menu.

## Fetching Objects

You can fetch previously stored items from a data bank or database. One of the common methods of working with data is to create a workfile and then fetch previously stored data into the workfile as needed.

To fetch objects into a workfile, select **Objects/Fetch from DB...** from the workfile menu or toolbar. You will see a dialog box prompting you for additional information for the fetch: objects to be fetched, directory and database location, as applicable.

See Chapter 6, “Fetching Objects from the Database” on page 106, for details on the advanced features of fetch.

## Updating Objects

Updating works like fetching objects, but requires that the objects be present in the workfile. To update objects in the workfile, select them from the workfile window, and click on **Objects/Update from DB...** from the workfile menu or toolbar. The Fetch dialog will open, but with the objects to be fetched already filled in. Simply specify the directory and database location and click **OK**.

The selected objects will be replaced by their counterparts in the data bank or database.

See Chapter 6 for additional details on the process of updating objects.

## Copy and Paste of Object Information

You can copy the list of object information displayed in a workfile or database window to the Windows clipboard and paste the list to other program files such as word processing files or spreadsheet files. Simply highlight the objects in the workfile directory window, select **Edit/Copy** (or click anywhere in the highlighted area, with the right mouse button, and select **Copy**). Then move to the application (word processor or spreadsheet) where you want to paste the list, and select **Edit/Paste**.

If only names are displayed in the window, EViews will copy a single line containing the highlighted names to the clipboard, with each name separated by a space. If the window contains additional information, either because **View/Display Comments (Label + -)** has been chosen in a workfile window or a query has been carried out in a database window, each name will be placed in a separate line along with the additional information.

Note that if you copy and paste the list of objects into another EViews workfile, the objects themselves will be copied.

## Commands

To create a new workfile, follow the `workfile` command with the name of the workfile. For example, if you type

```
workfile test1
```

EViews brings up the Workfile Range dialog to specify the range of the new workfile TEST1.

To save a workfile, follow the `save` command with a name for the saved workfile. For example,

```
save test2
```

saves the active workfile under the name TEST2 in the default directory.

See the *Command and Programming Reference* for a complete list of commands and options available in EViews.

## Chapter 4. Basic Data Handling

---

The process of entering, reading, editing, manipulating, and generating data forms the foundation of most data analyses. Accordingly, most of your time in EViews will probably be spent working with data. EViews provides you with a sophisticated set of data manipulation tools that make these tasks as simple and straightforward as possible.

This chapter describes the fundamentals of working with data in EViews. There are three cornerstones of data handling in EViews: the two most common data objects, *series* and *groups*, and the use of *samples*. We begin with a brief description of series, groups, and samples, and then discuss basic input, output, and editing of data.

In the next chapter, we discuss EViews' powerful language for generating and manipulating the data in series and groups.

### Data Objects

The actual numeric values that make up your data will generally be held in one or more of EViews' data objects (series, groups, matrices, vectors, scalars). For most users, series and groups will be, by far, the most important objects, so they will be the focus of our discussion. Matrices, vectors, and scalars are discussed at length in the *Command and Programming Reference*.

The following discussion is intended to provide only a brief introduction to the basics of series and groups. Our goal is to describe the basics of data input and output in EViews. An in-depth discussion of series and group objects follows in subsequent chapters.

### Series

An EViews series contains a set of observations on a variable. Associated with each observation in the series is a date or observation label. For series in dated workfiles, the observations are presumed to be observed regularly over time. For undated data, the observations are not assumed to follow any particular frequency.

#### Creating a series

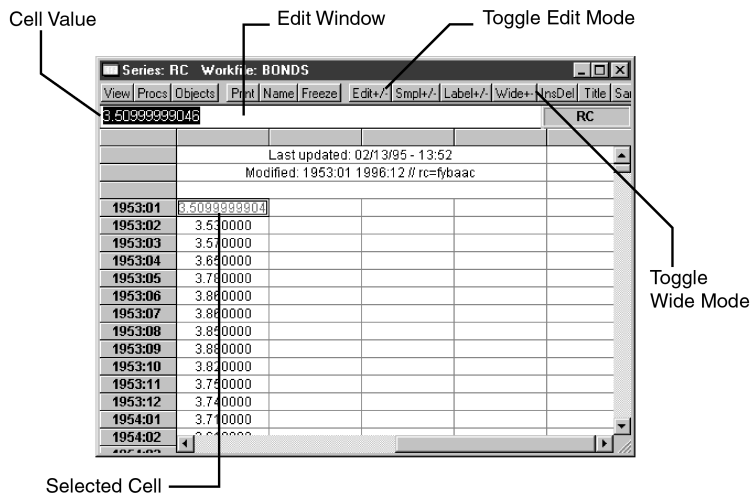
One method of creating a series is to select **Object/New Objects...** from the menu, and then to select **Series**. You may, at this time, provide a name for the series, or you can let the new series be untitled. Click **OK**. EViews will open a spreadsheet view of the new series object. All of the observations in the series will be assigned the missing value code "NA". You can then edit or use expressions to assign values for the series.

The second method of creating a series is to generate the series using mathematical expressions. Click on **Quick/Generate Series...** and enter the expression defining the series. We will discuss this method in depth in the next chapter.

## Editing a series

You can edit individual values of the data in a series.

- First open the spreadsheet view of the series. If the series window display does not show the spreadsheet view, click on the **Sheet** button, or select **View/Spreadsheet**, to change the default view.
- Next, make certain that the spreadsheet window is in *edit mode*. EViews provides you with the option of protecting the data in your series by turning off the ability to edit from the spreadsheet window. You can use the **Edit +/-** button on the toolbar to toggle between edit mode and *protected mode*:



Here we see an edit mode example for a series spreadsheet window. Notice the presence of the edit window just underneath the series toolbar containing the value of RC in 1953:01, and the double box around the selected cell in the spreadsheet—neither are present in protected mode.

- To change the value for an observation, select the cell, type in the value, and press ENTER. For example, to change the value of RC in 1953:01, simply click on the cell containing the value, type the new value in the edit window, and press ENTER.

Note that some cells are protected. If you select one of the protected cells, EViews will display a message in the edit window telling you that the cell cannot be edited.



- When you have finished editing, you should protect yourself from inadvertently changing values of your data by clicking on **Edit** +/- to turn off edit mode.

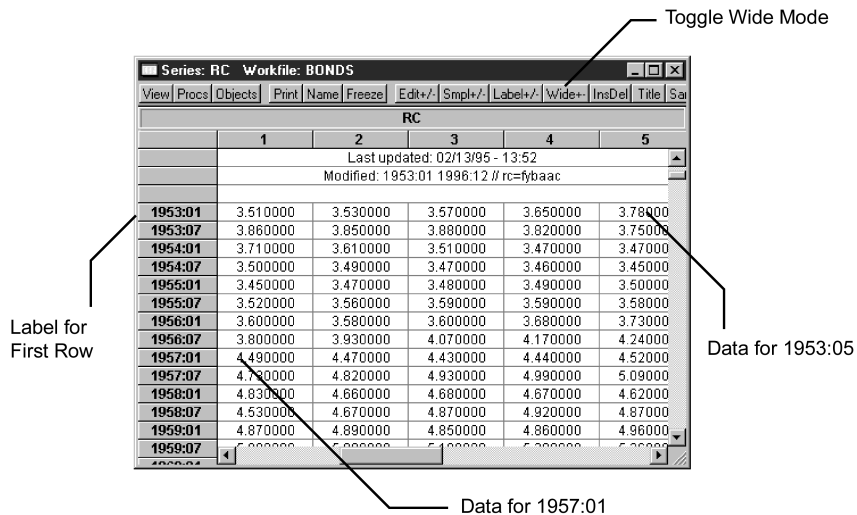
### Changing the Spreadsheet Display

EViews provides you with several different ways of looking at your data in spreadsheet form.

The *narrow display* displays the observations for the series in a single column, with date labels in the margin.

The *wide display* arranges the observations from left to right and top to bottom, with the label for the first observation in the row displayed in the margin. For dated workfiles, EViews will, if possible, arrange the data in a form which matches the frequency of the data. Thus, semi-annual data will be displayed with two observations per row, quarterly data will contain four observations per row, and monthly data will contain six observations in each row.

You can change the display to show the observations in your series in a single column by clicking on the **Wide** +/- button on the spreadsheet view toolbar (you may need to resize the series window to make this button visible). For example, toggling the **Wide** +/- button switches the display between the compact display depicted below, and the single column display:



This compact display format is useful when you wish to arrange the observations with data for a particular season displayed in the columns.

By default, all observations in the workfile are displayed, even those observations not in the current sample. By pressing **Smpl** +/- you can toggle between showing all observations in the workfile, and showing only those observations in the current sample.

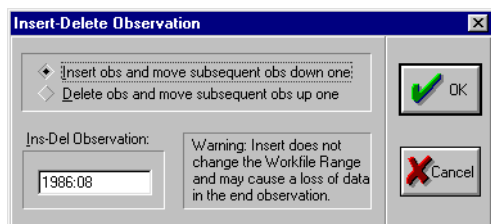
There are two features that you should keep in mind as you toggle between the various display settings:

- If you choose to display only the observations in the current sample, EViews will switch to single column display.
- If you switch to wide display, EViews automatically turns off the display filter so that all observations in the workfile are displayed.

One consequence of this behavior is that if you begin with a narrow display of observations in the current sample, click on **Wide** +/- to switch to wide display, and then press the **Wide** +/- button again, EViews will provide a narrow display of all of the observations in the workfile. To return to the original narrow display of the current sample, you will need to press the **Smpl** +/- button again.

### Inserting and deleting observations in a series

You can also insert and delete observations in the series. To insert an observation, first click on the cell where you want the new observation to appear. Next, click on the **InsDel** button. You will see a dialog asking whether you wish to insert or delete an observation at the current position:



You may change the location by editing the observation box. If you choose to insert an observation, EViews will insert a missing value at the appropriate position and push all of the observations down so that the last observation will be lost from the workfile. If you wish to preserve this observation, you will have to expand the workfile before inserting observations. If you choose to delete an observation, all of the remain-

ing observations will move up, so that you will have a missing value at the end of the workfile range.

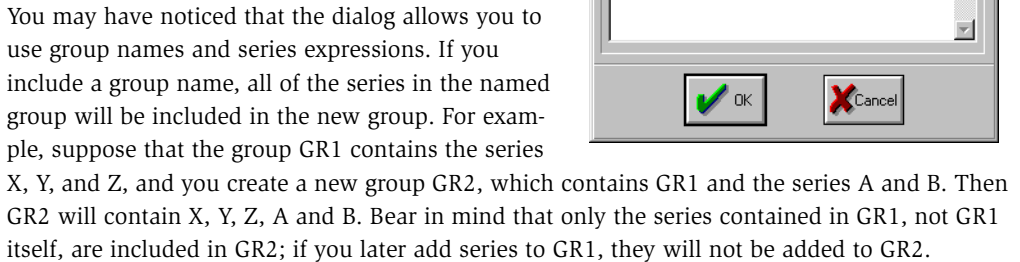
### Groups

When working with multiple series, you will often want to create a group object to help you manage your data. A group is a list of series names (and potentially, mathematical expressions) that provides simultaneous access to all of the elements in the list.

With a group, you can refer to sets of variables using a single name. Thus, a set of variables may be analyzed, graphed, or printed using the group object, rather than each one of the individual series. Therefore, groups are often used in place of entering a lengthy list of names. Once a group is defined, you can use the group name in many places to refer to all of the series contained in the group.

## Creating Groups

You should enter the names of the series to be included in the group, and then click **OK**. A group window will open showing a spreadsheet view of the group.



For example, here is a spreadsheet view of an untitled group containing the series RC, and series expressions for the lag of RG,  $RG(-1)$ , and a series expression involving RC and RG:

Group: UNTITLED    Workfile: BONDS			
View	Procs	Objects	Print   Name   Freeze   Edit+/-   Smpl+/-   InsDel   Transpose
obs	RC	RG(-1)	2*RC/RG
1953:01	3.510000		NA
1953:02	3.530000	0.000000	NA
1953:03	3.570000	0.000000	NA
1953:04	3.850000	0.000000	2.579505
1953:05	3.780000	2.830000	2.478689
1953:06	3.860000	3.050000	2.482315
1953:07	3.860000	3.110000	2.634812
1953:08	3.850000	2.930000	2.610169
1953:09	3.880000	2.950000	2.703833
1953:10	3.820000	2.870000	2.872180
1953:11	3.750000	2.660000	2.798507
1953:12	3.740000	2.680000	2.888031
1954:01	3.710000	2.590000	2.991935
1954:02			

You can also create an empty group that may be used for entering new data from

the keyboard or pasting data copied from another Windows program. This method is described in detail below in our discussion of entering data.

### Editing in a Group

Editing data in a group is similar to editing data in a series. Open the group window, and click on **Sheet**, if necessary, to display the spreadsheet view. If the group spreadsheet is in protected mode, click on **Edit** +/- to enable edit mode, then select your cell, enter the new value, and press RETURN. The new number should appear in the spreadsheet.

Since groups are simply references to series, editing the series within a group changes the values in the original series.

As with series spreadsheet views, you may click on **Smpl** +/- to toggle between showing all of the observations in the workfile and showing only those observations in the current sample. Unlike the series window, the group window always shows series in a single column.

## Samples

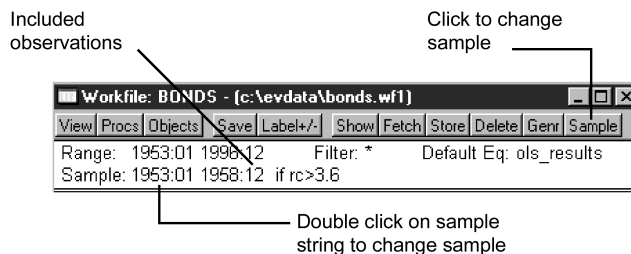
One of the most important concepts in EViews is the *sample* of observations. The sample is the set (often a subset) of observations in the workfile to be included in displays and in statistical procedures. Samples may be specified using ranges of observations and/or “if conditions” that observations must satisfy to be included.

For example, you can tell EViews that you want to work with observations from 1953:1 to 1970:12 and 1995:1 to 1996:12. Or you may want to work with data from 1953:1 to 1958:12 where observations on the series RC exceed 3.6.

### The Workfile Sample

When you create a workfile, the *global* or *workfile sample* is set initially to be the entire range of the workfile. The workfile sample tells EViews what set of observations you wish to use for subsequent operations. Unless you want to work with a different set of observations, you will not need to reset the workfile sample.

You can always tell the current workfile sample of observations by looking at the top of your workfile window:

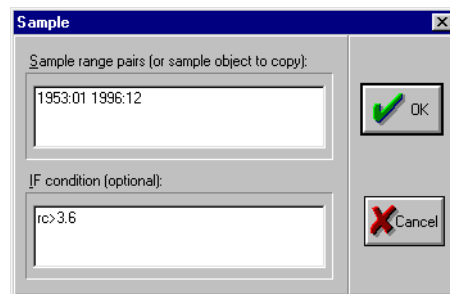


Here the sample consists of those observations between 1953:01 and 1958:12 for which the value of the RC exceeds 3.6.

## Changing the Sample

There are three ways to set the global sample: you can click on the **Sample** button in the workfile toolbar, you can select **Objects/Sample...** from the menu, or you could click on the sample string in the workfile window display.

In the upper edit window you will enter one or more pairs of dates. Each pair identifies a starting and ending observation for a set of observations to be included in the sample. For example, if you entered the string “1950 1980 1990 1995”, EViews will use observations for 1950 through 1980 and observations for 1990 through 1995 in subsequent operations; observations from 1981 through 1989 will be excluded.



EViews provides some special shortcuts that may make entering sample range pairs easier. First, you can use the keyword `@all`, to refer to the entire workfile range. In the workfile above, entering `@all` in the dialog is equivalent to entering `1953:1 1996:12`. Furthermore, you may use `@first` and `@last` to refer to the first and last observation in the workfile. Thus, all three of the following sample ranges are identical:

```
@all
@first 1996:12
1953:1 @last
```

Sample range elements may contain mathematical expressions to create date offsets. This feature can be particularly useful in setting up a fixed width window of observations. For example, in the workfile above, the sample strings

```
1953:1 1953:1+11
```

define a sample that includes the 12 observations in the calendar year beginning in 1953:1.

While EViews expects date offsets that are integer values, there is nothing to stop you from adding or subtracting non-integer values—EViews will automatically convert the number to an integer. You should be warned, however, that the conversion behavior is not guaranteed to be well-defined. If you must use non-integer values, you are strongly encouraged to use the `@round`, `@floor` or `@ceil` functions to enforce the desired behavior.

The lower part of the sample window allows you to add conditions to the sample specification. The sample is the intersection of the set of observations defined by the range pairs in the upper window and the set of observations defined by the if conditions in the lower window. For example, if you enter:

**Upper window:** 1980 1993

**Lower window:** `incm > 5000`

the sample includes observations for 1980 through 1993 where the series INCM is greater than 5000.

Similarly, if you enter:

**Upper window:** 1958:1 1998:1

**Lower window:** `gdp > gdp(-1)`

all observations between the first quarter of 1958 and the last quarter of 1998, where GDP has risen from the previous quarter, will be included.

The OR and AND operators allow for the construction of more complex expressions. For example, suppose you now wanted to include in your analysis only those individuals whose income exceeds 5000 dollars per year and who have at least 13 years of education. Then you can enter:

**Upper window:** `@all`

**Lower window:** `income > 5000 and educ > = 13`

Multiple range pairs and if conditions may also be specified:

**Upper window:** 50 100 200 250

**Lower window:** `income > = 4000 and educ > 12`

includes undated workfile observations 50 through 100 and 200 through 250, where the series INCOME is greater than or equal to 4000 and the series EDUC is greater than 12.

You can create even more elaborate selection rules by including EViews built-in functions:

**Upper window:** 1958:1 1998:1

**Lower window:** `(ed > = 6 and ed < = 13) or earn < @mean(earn)`

includes all observations where the value of the variable ED falls between 6 and 13, or where the value of the variable EARN is lower than its mean. Note that you may use parentheses to group the conditions and operators when there is potential ambiguity in the order of evaluation.

It is possible that one of the comparisons used in the conditioning statement will generate a missing value. For example, if an observation on INCM is missing, then the comparison  $\text{INCM} > 5000$  is not defined for that observation. EViews will treat such missing values as though the condition were false, and the observation will not be included in the sample.

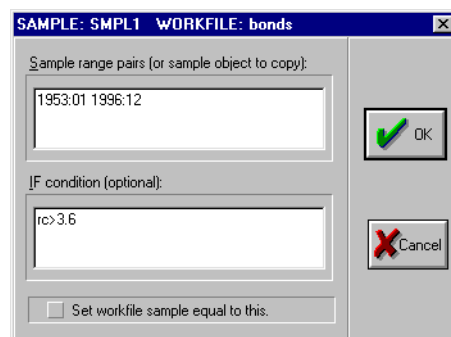
## Sample Objects

As you have seen, it is possible to develop quite elaborate selection rules for the workfile sample. However, it can quickly become cumbersome and time-consuming to re-enter these rules if you change samples frequently. Fortunately, EViews provides you with a method of saving sample information in a sample object which can then be referred to by name. If you work with various well-defined subsets of your data, you will soon find sample objects to be indispensable.

## Creating a Sample Object

To create a sample object, select **Objects/New...** When the New Objects dialog appears, select **Sample** and optionally, provide a name. If you do not provide a name, EViews will automatically assign one for you (sample objects may not be untitled). Click on **OK** and EViews will open a dialog:

Here is a partially filled in sample object dialog named SMPL1. Notice that while this dialog looks very similar to the one we described above for setting the sample, there are minor cosmetic differences: the name of the sample object appears in the title bar, and there is a check box for setting the workfile sample equal to this sample object.



These cosmetic differences reflect the two distinct purposes of the dialog: (1) to define the sample object, and (2) to set the workfile sample. Since EViews separates the act of defining the sample object from the act of setting the workfile sample, you can define the object without changing the workfile sample, and vice versa.

To define the sample object, you should fill out this dialog as before and click on **OK**. The sample object now appears in the workfile directory.

### Using a Sample Object

Once created, a sample object can be referred to by its name almost anywhere that EViews prompts you for a sample description.

Alternatively, you can use a previously defined sample object directly to set the workfile sample. Simply open a sample object by double clicking on the name or icon. This will reopen the dialog. If you wish to change the sample object, edit the sample specification; otherwise, simply click **Set workfile sample** on the check box and click on **OK**.

### Command Window Methods

You may find it easier to work with samples from the command window. Instead of specifying the sample by clicking on **Sample** or selecting **Quick/Sample...**, you can always enter the sample from the command window using the `smpl` command. Simply click on the command window to make it active, and type the keyword `smpl`, followed by the sample string:

```
smpl 1955:1 1958:12 if rc>3.6
```

and then press ENTER (notice, in the example above, the use of the keyword IF to separate the two parts of the sample specification). You should see the sample change in the workfile window.

If you wish to use sample objects, you will first create the sample object using a `sample` declaration, followed by the name to be given to the sample object, then the sample string:

```
sample mysample 1955:1 1958:12 if rc>3.6
```

To set the workfile sample using the sample object, just enter the `SMPL` command, followed by the sample object name. For example,

```
smpl mysample
```

will set the workfile sample according to the rules contained in the sample object `MYSAMPLE`.

## Importing Data

The data for your project may be available in a variety of forms. Perhaps they are in an external database such as the DRI Basic Economics, or a Haver Analytics database. The data may also be in a machine readable spreadsheet or text file that you created yourself or downloaded from the World Wide Web, or perhaps they are in a book or photocopy form.

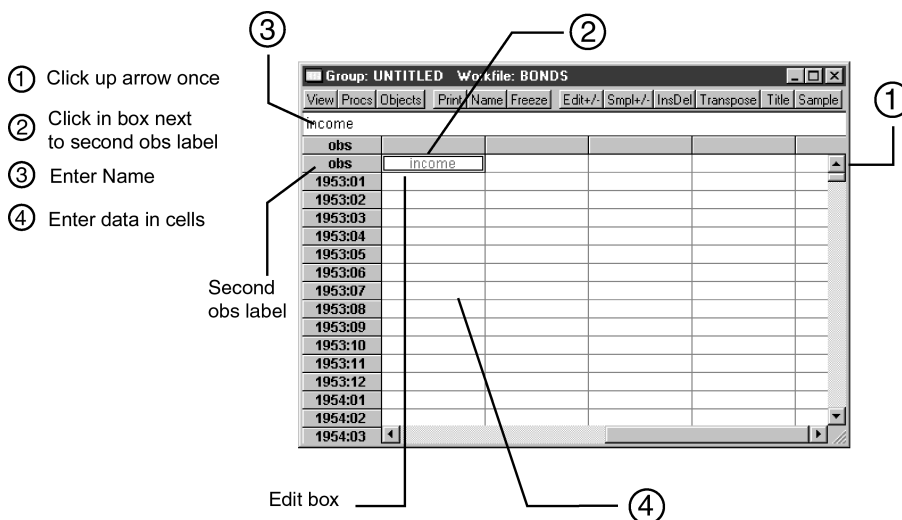
There are several methods for getting data into EViews. In the discussion below, we outline the basics of data input into series and group objects from spreadsheet, text file, or printed formats. We defer until later chapters a full discussion of working with workfiles and databases. We also omit, for the moment, discussion of input into EViews matrix, vector and pool objects.



## Entering Data

For small datasets in printed form, you may wish to enter the data by typing at the keyboard.

- Your first step is to open a temporary spreadsheet window in which you will enter the data. Choose **Quick/Empty Group (Edit Series)** from the main menu to open an untitled group window:



- The next step is to create and name the series. First click once on the up arrow to display the second **obs** label on the left-hand column. The row of cells next to the second **obs** label is where you will enter and edit series names.

Click once in the cell next to the second **obs** label, and enter your first series name. Here we have typed INCOME in the command window (the name in the cell changes as we type in the command window). Press RETURN.

- Repeat this procedure in subsequent columns for each additional series.

If you decide you want to rename one of your series, simply select the cell containing the series name, edit the name, and then press RETURN. EViews will prompt you to confirm the series rename.

- To enter the data, click on the appropriate cell and type the number. Pressing RETURN after entering a number will move you to the next cell. If you prefer, you can use the cursor keys to navigate the spreadsheet.
- When you are finished entering data, close the group window. If you wish, you can first name the untitled group by clicking on the **Name** button. If you do not wish to keep the group, answer **Yes** when EViews asks you to confirm the deletion.

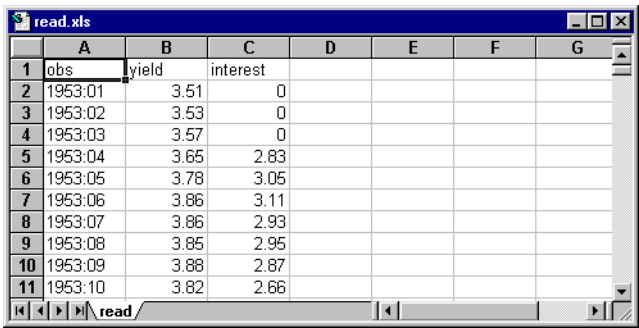
Copying and Pasting

The Windows clipboard is a handy way to move data within EViews and between EViews and other software applications. It is a natural tool for importing data from Excel and other Windows applications that support Windows copy-and-paste.

Copying from Windows applications

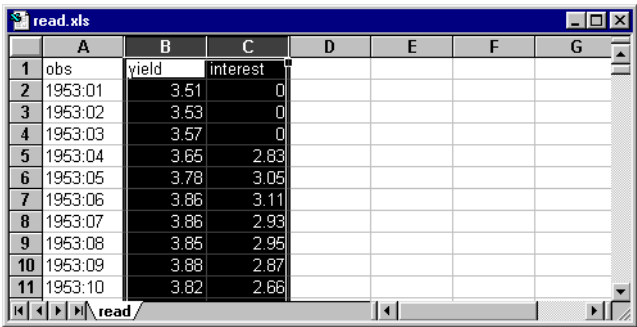
The following discussion involves an example using an Excel spreadsheet, but the basic principles apply for other Windows applications.

Suppose you have bond yield and interest rate data in an Excel spreadsheet that you would like to bring into EViews. Open the spreadsheet in Excel:



	A	B	C	D	E	F	G
1	obs	yield	interest				
2	1953:01	3.51	0				
3	1953:02	3.53	0				
4	1953:03	3.57	0				
5	1953:04	3.65	2.83				
6	1953:05	3.78	3.05				
7	1953:06	3.86	3.11				
8	1953:07	3.86	2.93				
9	1953:08	3.85	2.95				
10	1953:09	3.88	2.87				
11	1953:10	3.82	2.66				

Your first step is to highlight the cells to be imported into EViews. Since the column headings YIELD and INTEREST will be used as EViews variable names, you should highlight them as well. Since EViews understands dated data, and we are going to create a monthly workfile, you do not need to copy the date column. Instead, click on the column label B and drag to the column label C. The two columns of the spreadsheet will be highlighted:



	A	B	C	D	E	F	G
1	obs	yield	interest				
2	1953:01	3.51	0				
3	1953:02	3.53	0				
4	1953:03	3.57	0				
5	1953:04	3.65	2.83				
6	1953:05	3.78	3.05				
7	1953:06	3.86	3.11				
8	1953:07	3.86	2.93				
9	1953:08	3.85	2.95				
10	1953:09	3.88	2.87				
11	1953:10	3.82	2.66				

Select **Edit/Copy** to copy the highlighted data to the clipboard.

Pasting into New Series

Start EViews and create a new, or load an existing, monthly workfile containing the dates in the Excel spreadsheet (in our example 1953:1 through 1994:11). Make certain that the sample is set to include the same observations that you have copied onto the clipboard.

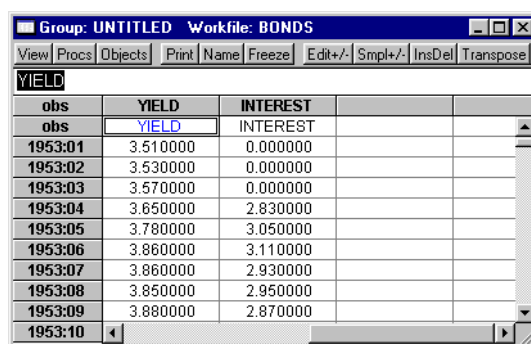
Select **Quick/Empty Group (Edit Series)**. Note that the spreadsheet opens in edit mode so there is no need to click the **Edit** +/- button.

Here, we have created a monthly workfile with a range from 1953:1 to 1999:12. The first row of the EViews spreadsheet is labeled 1953:01. Since we are pasting in the series names, you should click on the up arrow in the scroll bar to make room for the series names.

Place the cursor in the upper-left cell, just to the right of the second obs label. Then select **Edit/Paste** from the main menu (*not* **Edit** +/- in the toolbar). The group spreadsheet will now contain the data from the clipboard.

Here, the series YIELD and INTEREST have been created in the workfile and can be used in any EViews procedure. You may now close the group window and delete the untitled group without losing the two series.

Note that when importing data from the clipboard, EViews follows the Windows standard of tab-delimited free-format data with one observation per line. Since different applications use different whitespace and delimiter characters, attempting to cut-and-paste from nonstandard applications may produce unanticipated results.



obs	YIELD	INTEREST
1953:01	3.510000	0.000000
1953:02	3.530000	0.000000
1953:03	3.570000	0.000000
1953:04	3.650000	2.830000
1953:05	3.780000	3.050000
1953:06	3.860000	3.110000
1953:07	3.860000	2.930000
1953:08	3.850000	2.950000
1953:09	3.880000	2.870000
1953:10		

## Pasting into Existing Series

You can import data from the clipboard into an existing EViews series or group spreadsheet by using **Edit/Paste** in the same fashion. There are only a few additional issues to consider.

- To paste several series, you will first open a group window containing the existing series. The easiest way to do this is to click on **Show**, and then type the series names in the order they appear on the clipboard. Alternatively, you can create an untitled group by selecting the first series, click selecting each subsequent series (in order), and then double clicking to open.
- Next, make certain that the group window is in edit mode. If not, press the **Edit** +/- button to toggle between edit mode and protected mode. Place the cursor in the target cell, and select **Edit/Paste**.
- Finally, click on **Edit** +/- to return to protected mode.
- If you are pasting into a single series you will need to make certain that the series window is in edit mode, and that the series is viewed in a single column. If the series is in multiple col-

umns, push on the **Smpl** +/- button. **Edit/Paste** the data and click on **Edit** +/- to protect the data.

## Importing Data from a File

You can also read data directly from files created by other programs. Data may be in standard ASCII form or in either Lotus (.WKS, .WK1 or .WK3) or Excel (.XLS) spreadsheet formats.

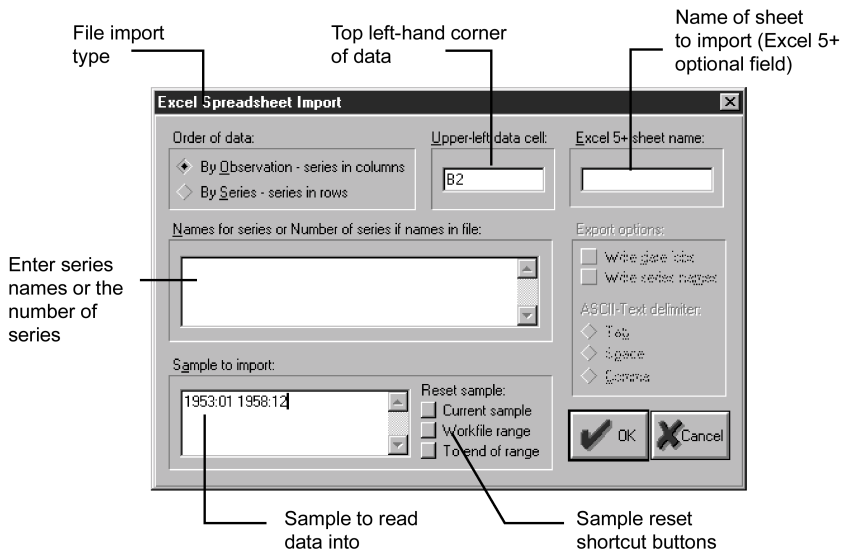
First make certain that you have an open workfile to receive the contents of the data import.

Next, click on **Procs/Import/Read Text-Lotus-Excel...** You will see a standard File dialog box asking you to specify the type and name of the file. Select a file type, navigate to the directory containing the file, and double-click on the name. Alternatively, type in the name of the file that you wish to read (with full path information, if appropriate); if possible, EViews will automatically set the file type, otherwise it will treat the file as an ASCII file. Click on **Open**.

EViews will open a dialog prompting you for additional information about the import procedure. The dialog will differ greatly depending on whether the source file is a spreadsheet or an ASCII file.

## Spreadsheet Import

The title bar of the dialog will identify the type of file that you have asked EViews to read. Here is the dialog for importing an Excel 5 (or later versions of Excel), spreadsheet:



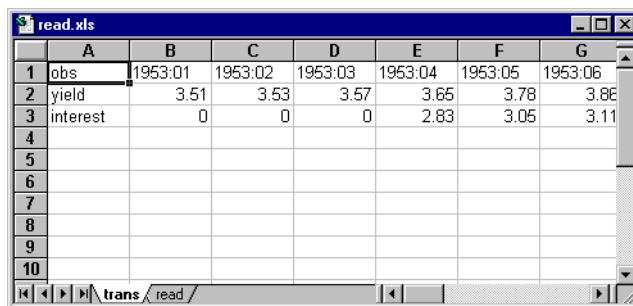
You will see slightly different versions of this dialog if you are reading a Lotus, or an Excel 4 (and earlier) file.

Now fill in the dialog:

- First, you need to tell whether the data are ordered by observation or by series. **By observation** means that all of the data for the first observation are followed by all of the data for the second observation, etc. **By series** means that all of the data for the first variable are followed by all data for the second variable, etc. Another interpretation for “by observation” is that variables are arranged in columns while “by row” implies that all of the observations for a variable are in a single row.

Our Excel example above is organized by observation since each series is in a separate column. If the Excel data for YIELD and INTEREST were each contained in a single row as depicted below, then the data should be read by series.

- Next, tell EViews the location of the beginning cell (upper left-hand corner) of your actual data, not including any label or date information. In both examples above, the upper left-hand cell is B2.



	A	B	C	D	E	F	G
1	obs	1953:01	1953:02	1953:03	1953:04	1953:05	1953:06
2	yield	3.51	3.53	3.57	3.65	3.78	3.86
3	interest	0	0	0	2.83	3.05	3.11
4							
5							
6							
7							
8							
9							
10							

- Enter the names of the series that you wish to read into the edit box. EViews reads spreadsheet data in contiguous blocks, so you should provide a name for each column or row (depending on the orientation of the data), even if you only wish to read selected rows.

Alternatively, if the names that you wish to use for your series are contained in the file, you can simply provide the number of series to be read. The names must be adjacent to your data. If the data are organized by row and the starting cell is B2, then the names must be in column A, beginning at cell A2. If the data are organized by column beginning in B2, then the names must be in row 1, starting in cell B1. If, in the course of reading the data, EViews encounters an invalid cell name, it will automatically assign the next unused name with the prefix SER, followed by a number (e.g., SER01, SER02, etc.).

- Lastly, you should tell EViews the sample of data that you wish to import. EViews begins with the first observation in the file and assigns it to the first date in the sample for each variable. Each successive observation in the file is associated with successive observations in the sample. Thus, in an annual workfile, if you enter the sample:

1971 1975 1990 1991

in the import dialog, the first 5 observations will be assigned to the dates 1971–1975, and the sixth and seventh observations will be assigned to the dates 1990–1991. The data in the intermediate period will be unaffected by the importing procedure.

You should be warned that if you read into a sample which has more observations than are present in your input file, observations for which there are no corresponding inputs will be assigned missing values. For example, if you read into the sample defined as 1971 1990, and there are only 10 observations in the input file, observations from 1981 to 1990 will be assigned missing values.

When the dialog is first displayed, EViews by default enters the current workfile sample in the edit box. You should edit this string to reflect the desired sample. To make setting the sample easier, EViews provides you with three push-buttons which change the string in the edit box to commonly used values:

**Current sample**—sets the dialog string to the current workfile sample.

**Workfile range**—sets the dialog string to the entire range of the workfile.

**To end of range**—sets the dialog string to all observations from the beginning of the current sample to the end of the workfile range.

- If you are reading data from an Excel 5 workbook file, there will be an additional edit box where you can enter the name of the sheet containing your data. If you do not enter a name, EViews will read the topmost sheet in the Excel workbook.
- When the dialog is completely filled out, simply click **OK** and EViews will read your file, creating series and assigning values as requested.

### ASCII Import

If you choose to read from an ASCII file, EViews will open an ASCII Text Import dialog. Fill out the dialog to read from the specified file.

The dialog box for ASCII file import is considerably more complicated than the corresponding spreadsheet dialog. While unfortunate, this complexity is necessary since there is no standard format for ASCII files. EViews provides you with a range of options to handle various types of ASCII files.

ASCII file importing is explained in considerable detail below in “Addendum: Reading ASCII Files”, beginning on page 70.

## Exporting Data

EViews provides you with a number of methods for getting data from EViews into other applications.

### Copying and Pasting

You can click and drag in a spreadsheet view or table of statistical results to highlight the cells you want to copy. Then click **Edit/Copy...** to put the data into the clipboard. You will see a dia-

log box asking whether to copy the numbers with the precision showing on your screen (formatted copy) or to copy the numbers at full precision (unformatted copy).

As a shortcut, you can highlight entire rows or columns of cells by clicking on the gray border that surrounds the spreadsheet. Dragging across the border selects multiple rows or columns. To copy several adjacent series from the spreadsheet, drag across their names in the top border. All of their data will be highlighted. Then click **Edit/Copy** to put the data into the clipboard

Once the data are on the clipboard, switch to the target application, and select **Edit/Paste**.

## Exporting to a File

First, click on **Procs/Export/Write Text-Lotus-Excel...**, then enter the name and type of the output file in the file dialog. As you fill out the file dialog, keep in mind the following behavior:

- If you enter a file name with an extension, EViews will use the file extension to identify the file type. Files with common spreadsheet extensions (.XLS, .WK3, .WK1, and .WKS) will be saved to the appropriate spreadsheet type. All others will be saved as ASCII files.
- If you do not enter an extension, EViews will use the file type selected in the combo-box to determine the output type. Spreadsheet files will have the appropriate extensions appended to the name. ASCII files will be saved in the name provided in the dialog, without an extension. EViews will not append extensions to ASCII files unless you explicitly include one in the file name.

Once you have specified the output file, click **OK** to open the export dialog.

Tip: if you highlight the series you wish to export before beginning the export procedure, the series names will be used to fill out the export dialog.

## Spreadsheet Export

The dialogs for spreadsheet export are virtually identical to the dialogs for spreadsheet import. You should determine the orientation of your data, the series to export, and the sample of observations to be written.

Additionally, EViews provides you with checkboxes for determining whether to include the series names and/or the series dates in the spreadsheet. If you choose to write one or both to the spreadsheet, make certain that the starting cell for your data leaves the necessary room along the borders for the information. If the necessary room is not available, EViews will ignore the option—for example, if you choose to write your data beginning in cell A1, EViews will not write the names or dates.

## ASCII Export

The ASCII export dialog is quite similar to the spreadsheet export dialog, but it contains a few additional options:

- You can change the text string to be used for writing missing values. Simply enter the text string in the edit field.
- EViews provides you with the option of separating data values with a tab, a space, or a comma. Click on the desired radio button.

You should be warned that if you attempt to write your data by series, EViews will write all of the observations for a series on a single line. If you have a reasonably long series of observations, these data may overflow the line-length of other programs.

## Commands

To create a new series from an existing series, follow the `series` or `genr` command with a name for the new series, an equals sign and an expression involving the existing series:

```
series logy = log(y)
```

creates a new series named LOGY that is the natural log of the series Y.

To create a new group of series, follow the `group` command with a name for the group and a list of series to include in the group, each separated by a space:

```
group rhs c x1 x2 z
```

creates a group named RHS that contains the constant C (a series of ones) and the series X1, X2, Z.

To view the series or group, follow the `show` command with the name of the series or group:

```
show logy
```

To bring up the import dialog box, follow the `read` command with the full name of the file (including the file extension) to import from:

```
read c:\data\cps88.dat
```

To bring up the export dialog box, follow the `export` command with the full name of the file (including the file extension) to export:

```
write a:\usmacro.dat
```

See the *Command and Programming Reference* for a complete list of commands and options available in EViews.

## Addendum: Reading ASCII Files

If you instruct EViews to read from an ASCII file, the ASCII Text Import dialog will be displayed.



You may notice that the dialog is more complicated than the corresponding spreadsheet dialog. Since there is no standard format for ASCII files we need to provide a variety of options to handle various types of files.

Note that the preview window at the bottom of the dialog shows you

the first 16K of your file. You can use this information to set the various formatting options in the dialog.

**ASCII Text Import**

Names for series or Number of series if names in file:

Data order:

- ☒ in Columns
- ☐ in Rows

Rectangular file layout options:

- ☒ File laid out as rectangle
- Columns to skip:
- Rows to skip:
- Comment character:

Series headers:

# of headers (including names) before data:

Sample to import:

1954:01 1994:12

Reset sample:

- ☐ Current sample
- ☐ Workfile range
- ☐ To end of range

Delimiters:

- ☒ Treat multiple delimiters as one
- ☐ Tab
- ☐ Comma
- ☒ Space
- ☐ Alpha (A-Z)
- ☐ Custom

Misc options:

- ☐ Quote with single ' not "
- ☐ Drop strings - don't make NA
- ☐ Numbers in (...) are negative
- ☐ Allow commas in numbers
- Currency:
- Text for NA:

Preview - First 16K of file:

LABEL	P85	P75	RMT85	CS82	SS82	S82	ME84	REV84	REG	CL
1	33	27	288	13	24	49	2135	2836	1	1
2	19	15	139	14	12	41	957	2035	1	1
3	26	20	196	12	14	41	1530	6030	1	1
4	19	15	159	12	19	41	1059	4704	1	1
5	56	52	536	20	27	61	3951	5183	1	1

OK Cancel

You must provide the following information:

- **Names for series or Number of series if names in file.** If the file does not contain series names, or if you do not want to use the names in the file, list the names of the series in the order they appear in the file, separated by spaces.

If the names of the series are located in the file *before the start of the data*, you can tell EViews to use these names by entering a number representing the number of series to be read.

If possible, you should avoid using parentheses and mathematical symbols such as \*, +, -, /, ^ in the names. If EViews tries to read the names from the file and encounters an invalid name, it will try to rename the series to a valid name by replacing invalid characters with underscores and numbers. For example, if the series is named X(-3) in the file, EViews will rename this series to X\_\_3\_01. If X\_\_3\_01 is already a series name, then EViews will name the series X\_\_3\_02, and so forth.

If EViews cannot name your series, say, because the name is a reserved name, or because the name is used by an object that is not a series, the series will be named SER01, SER02, etc.

You should be very careful in naming your series and listing the names in the dialog. If the name in the list or in the file is the same as an existing series name in the workfile, the data in the existing series will be overwritten.

- **Data order.** You need to specify how the data are organized in your file. If your data are ordered by observation so that each series is in a column, select **in Columns**. If your data are ordered by series so that all the data for the first series are followed by all the data for the second series and so on, select **in Rows**.
- **Sample to import.** You should specify the sample in which to place the data from the file. EViews fills out the dialog with the current workfile sample, but you can edit the sample string or use the sample reset buttons to change the input sample. The input sample only sets the sample for the import procedure, it does not alter the workfile sample.

EViews fills all of the observations in the current sample using the data in the input file. There are a couple of rules to keep in mind:

1. EViews assigns values to all observations in the input sample. Observations outside of the input sample will not be changed.
2. If there are too few values in the input file, EViews will assign NAs to the extra observations in the sample.
3. Once all of the data for the sample have been read, the remainder of the input file will be ignored.

In addition to the above information, you can use the following options to further control the way EViews reads in ASCII data.

EViews scans the first few lines of the source file and sets the default formatting options in the dialog based on what it finds. However, these settings are based on a limited number of lines and may not be appropriate. You may find that you need to reset these options.

## Delimiters

Delimiters are the characters that your file uses to separate observations. You can specify multiple delimiters by selecting the appropriate entries. **Tab**, **Comma**, and **Space** are self-explanatory. The **Alpha** option treats any of the 26 characters from the alphabet as a delimiter.

For delimiters not listed in the option list, you can select the **Custom** option and specify the symbol you wish to treat as a delimiter. For example, you can treat the slash “/” as a delimiter by selecting **Custom** and entering the character in the edit box. If you enter more than one character, *each* character will be treated as a delimiter. For example, if you type “//” in the **Custom** field, then the single slash “/” will be treated as a delimiter, instead of the double slash “//”. The double slash will be interpreted as two delimiters.

EViews provides you with the option of treating multiple delimiter characters as a single delimiter. For example, if “,” is a delimiter and the option **Treat multiple delimiters as one** is selected, EViews will interpret “,,,” as a single delimiter. If the option is turned off, EViews will view this string as two delimiters surrounding a missing value.

## Rectangular File Layout Options

To treat the ASCII file as a rectangular file, select the **File laid out as rectangle** option in the upper right-hand portion of the dialog. If the file is rectangular, EViews reads the file as a set of *lines*, with each new line denoting a new observation or a new series. If you turn off the rectangular option, EViews treats the whole file as one long string separated by delimiters and carriage returns.

Knowing that a file is rectangular makes ASCII reading much simpler since EViews knows how many values to expect on a given line. For files that are not rectangular, you will need to be precise about the number of series or observations that are in your file. For example, suppose that you have a non-rectangular file that is ordered in columns and you tell EViews that there are four series in the file. EViews will ignore new lines and will read a new observation after reading every four values.

If the file is rectangular, you can tell EViews to skip columns and/or rows.

For example, if you have a rectangular file and you type 3 in the **Rows to skip** field, EViews will skip the first three rows of the data file. Note that you can only skip the *first* few rows or columns; you cannot skip rows or columns in the middle of the file.

## Series Headers

This option tells EViews how many “cells” to offset as series name headers before reading the data in the file. The way that cell offsets are counted differs depending on whether the file is in rectangular form or not.

For files in rectangular form, the offsets are given by rows (for data in columns) or by columns (for data in rows). For example, suppose your data file looks as follows:

	LABEL	P85	P75	RMT85	CS82	SS82	S82	ME84	REV84	REG	CL
1	33	27	288	13	24	49	2135	2836	1	1	
2	19	15	139	14	12	41	957	2035	1	1	
3	26	20	196	12	14	41	1530	6030	1	1	

There is a one line (row) gap between the series name line and the data for the first observation. In this case, you should set the series header offset as 2, one for the series name line and one for the gap. If there were no gap, then the offset should be specified as 1.

For files that are in non-rectangular form, the offsets are given by the number of cells separated by the delimiters. For example, suppose you have a data file that looks as follows:

FTP	UEMP	MAN	LIC	GR	CLEAR	WM	NMAN	GOV	HE
WE	HOM	ACC	ASR						
260.35	11.0	455.5	178.15	215.98	93.4	558724.	538.1	133.9	2.98
117.18	8.60	39.17	306.18						
269.80	7.0	480.2	156.41	180.48	88.5	538584.	547.6	137.6	3.09
134.02	8.90	40.27	315.16						

The data are ordered in columns but each observation is recorded in two lines, the first line for the first 10 series and the second line for the remaining 4 series.

It is instructive to examine what happens if you incorrectly read this file as a rectangular file with 14 series and a header offset of 2. EViews will look for the series names in the first line, will skip the second line, and will begin reading data starting with the third line, treating each line as one observation. The first 10 series names will be read correctly, but since EViews will be unable to find the remaining four names on the first line, the remaining series will be named SER01–SER04. The data will also be read incorrectly. For example, the first four observations for the series GR will be 215.9800, NA, 180.4800, and NA, since EViews treats each line as a new observation.

To read this data file properly, you should turn off the rectangle file option and set the header offset to 1. Then EViews will read, from left to right, the first 14 values that are separated by a delimiter or carriage return and take them as series names. This corresponds to the offset of 1 for headers. The next 14 observations are the first observations of the 14 series, and so on.

## Misc. Options

**Quote with single ‘ not “.** The default behavior in EViews is to treat anything inside a pair of matching double quotes as one string, unless it is a number. This option treats anything inside a pair of matching single quotes as one string, instead of the double quotes. Since EViews does not support strings, the occurrence of a pair of matching double quotes will be treated as missing, unless the thing inside the pair of double quotes is a number.

**Drop strings—don’t make NA.** Since EViews does not support strings, any input into a series observation that is not a number or delimiter will, by default, be treated as a missing observation. For example, 10b and 90:4 will both be treated as missing values (unless Alphabetic characters or “:” are treated as delimiters). The **Drop strings** option will skip these strings instead of treating them as NAs.

If you choose this option, the series names, which are strings, will also be skipped so that your series will be named using the EViews default names: SER01, SER02, and so on. If you wish to name your series, you should list the series names in the dialog.

Note that strings that are specified as missing observations in the **Text for NA** edit box (see below) will not be skipped and will be properly indicated as missing.

**Numbers in ( ) are negative.** By default, EViews treats parentheses as strings. However, if you choose this option, numbers in parentheses will be treated as negative numbers and will be read accordingly.

**Allow commas in numbers.** By default, commas are treated as strings unless you specify them as a delimiter. For example, 1,000 will be read as either NA (unless you choose the drop string option, in which case it will be skipped) or as two observations, 1 and 0 (if the comma is a delimiter). However, if you choose to **Allow commas in numbers**, 1,000 will be read as the number 1000.

**Currency.** This option allows you to specify a symbol for currency. For example, the default behavior treats \$10 as a string (which will either be NA or skipped) unless you specify \$ as a delimiter. If you enter “\$” in the **Currency** option field, then \$10 will be read as the number 10.

The currency symbol can appear either at the beginning or end of a number but not in the middle. If you type more than one symbol in the field, each symbol will be treated as a currency code. Note that currency symbols are case sensitive. For example, if the Japanese yen is denoted by the “Y” prefix, you should enter “Y”, not “y”.

**Text for NA.** This option allows you to specify a code for missing observations. The default is NA. You can use this option to read data files that use special values to indicate missing values, e.g., “.”, or “-99”.

You can specify only one code for missing observations. The entire **Text for NA** string will be treated as the missing value code.

## Examples

In these examples, we demonstrate the ASCII import options using example data files downloaded from the World Wide Web. The first example file looks as follows:

X	Y	Z	A	B	AA	BB
5	3	10	-0.6008	NBA	10	8
10	3	7	-0.4837	NFL	7	2
2	2	5	1.2467	NFL	0	0
12	0	3	0.1705	NBA	5	1

This is a cross-section data set, seven series ordered in columns, each separated by a single space. Note that the B series takes string values, which will be replaced by NAs. If we type 7 series in the number of series field and use the default setting, EViews will correctly read in the data.

By default, EViews checks the **Treat multiple delimiters as one** option even though the series are delimited by a single space. If you do not check this option, the last series BB will not be read in and EViews creates a series named SER01. This strange behavior is caused by an extra space in the very first column of the data file, before the 1st and 3rd observations of the X series. EViews treats the very first space as delimiter and looks for the first series data before the first extra space, which is missing. Therefore the first series is named SER01 with data NA, 10, NA, 12 and all other series are incorrectly imported.

To handle this case, EViews automatically ignores the delimiter before the first column data if you choose both the **Treat multiple delimiters as one** and the **File laid out as rectangle** options.

The top of the second example file looks like:

African elephant	6654.000	5712.000	-999.0	-999.0	3.3	38.6	645.0
African giant pouched rat	1.000	6.600	6.3	2.0	8.3	4.5	42.0
Arctic Fox	3.385	44.500	-999.0	-999.0	12.5	14.0	60.0
Arctic ground squirrel	.920	5.700	-999.0	-999.0	16.5	-999.0	25.0
Asian elephant	2547.000	4603.000	2.1	1.8	3.9	69.0	624.0
Baboon	10.550	179.500	9.1	.7	9.8	27.0	180.0

This is a cross-section data set, ordered in columns, with missing values coded as -999.0. There are eight series, each separated by spaces. The first series is the ID name in strings.

If we use the EViews defaults, there will be problems reading this file. The spaces in the ID description will generate spurious NA values in each row, breaking the rectangular format of the file. For example, the first name will generate two NAs, since “African” is treated as one string, and “elephant” as another string.

You will need to use the **Drop strings** option to skip all of the strings in your data so that you don’t generate NAs. Fill out the ASCII dialog as follows:

Note the following:

- Since we skip the first string series, we list only the remaining seven series names.
- There are no header lines in the file so we set the offset to 0.
- If you are not sure whether the delimiter is a space or tab, mark both options. You should treat multiple delimiters as one.
- **Text for NA** should be entered exactly as it appears in the file. For this example you should enter -999.0, not -999.

The third example is a daily data file that looks as follows:

```

"Daily Corporate Bond Yields"
"Jan. 1, 1989 to Dec. 31, 1993"

"Source: Federal Reserve Statistical Release H.15"
"Board of Governors of the Federal Reserve System"

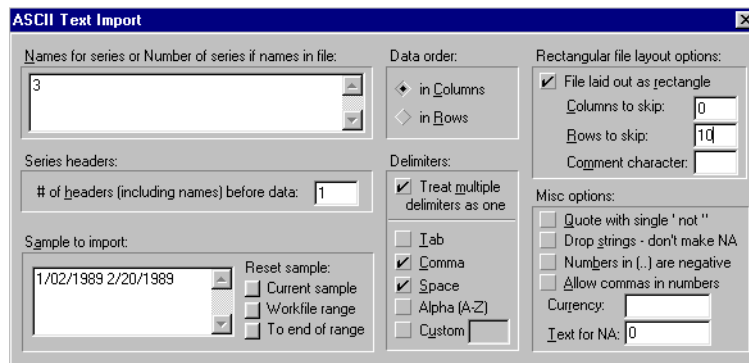
"DATE      = MM/DD/YY FORMAT"
"AA        = MOODY'S SEASONED AA"
"BAA       = MOODY'S SEASONED BAA"

"DATE",      "AA",      "BAA"
01/02/89      0          0
01/03/89      9.88       10.71
01/04/89      9.86       10.71
01/05/89      9.89       10.75

```

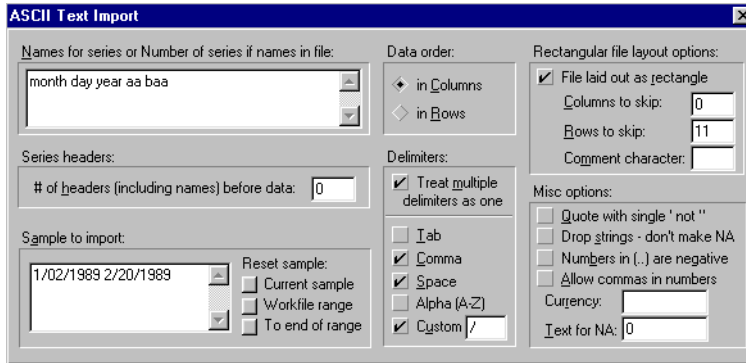
This file has 10 lines of data description, line 11 is the series name header, and the data begin in line 12. The data are ordered in columns in rectangular form with missing values coded as 0.

The simplest approach to reading this file is to set the options as follows:



Here you skip the first 10 rows of the rectangular file, and read three series with the names in the file. The only problem with this method is that the DATE series will be filled with NAs since EViews treats the entry as a string (because of the “/” in the date entry).

You can avoid this problem by identifying the slash as a delimiter using the **Custom** edit box. The first column will now be read as three distinct series since the two slashes are treated as delimiters. Therefore, we modify the option settings as follows:



Note the changes to the dialog entries:

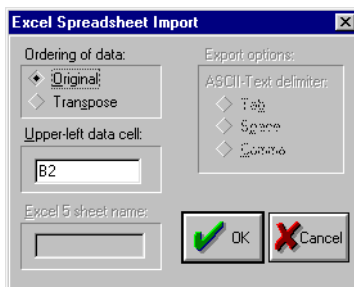
- We now list five series names. We cannot use the file header since the line only contains three names.
- We skip 11 rows with no header offset since we want to skip the name header line.
- We specify the slash “/” as an additional delimiter in the **Custom** option field.

The month, day, and year will be read as separate series and can be used as a quick check of whether the data have been correctly read.

## Addendum: Matrix Object Reading and Writing

The two available procedures for matrices allow you to import and export data directly to or from the matrix object. As with the standard EViews procedures, you can read or write from spreadsheet or from ASCII files.

Import Data (ASCII,XLS,WK?)...  
Export Data (ASCII,XLS,WK?)...



To read from a file, select **Procs/Import Data (ASCII, .XLS, .WK?)...** EViews will open an import dialog.

Here, we depict the matrix import Excel dialog. As above, the corresponding ASCII dialog has many more options, since ASCII file reading is more complicated. Note that both the import and export dialogs differ little from the series import dialogs described above. The differences reflect the different nature of series and matrix input and output. For example, dialog options for series names and the sample are omitted since they do not apply to matrices.

ces.



To write the contents of the matrix to a file, select **Procs/Export Data (ASCII, .XLS, .WK?)...** and fill in the dialog.

In reading from a file, EViews first fills the matrix with NAs, puts the first data element in the (1,1) element of the matrix, and then continues reading the data by row or column according to the specified settings for **Ordering of data**. If this option is set as **Original**, EViews will read by row, filling the first row from left to right, and then continuing on to the next row. If the ordering is set as **Transpose**, EViews fills the matrix by columns, reading the first column from top to bottom and then continuing on to the next column.

ASCII files provide you with the option of reading your file as a rectangle. If your ASCII file is laid out as a rectangle, the contents of the rectangle will be placed in the matrix beginning at the (1,1) element of the matrix. For example, if you have a  $3 \times 3$  matrix X, and read from the ASCII file containing:

```
1 2 3 4
5 6 7 8
9 10 11 12
```

using the **File laid out as rectangle** option, the matrix X will contain the corresponding rectangular portion of the ASCII file:

```
1 2 3
5 6 7
9 10 11
```

If you do not select the rectangular read option, EViews fills the matrix element-by-element, reading from the file line by line. Then X will contain:

```
1 2 3
4 5 6
7 8 9
```



## Chapter 5. Working with Data

---

This chapter describes EViews' powerful language for using expressions and generating and manipulating the data in series and groups. We first describe the fundamental rules for working with mathematical expressions in EViews, and then describe how to use these expressions in working with series and group data.

### Using Expressions

One of the most powerful features of EViews is the ability to use and to process mathematical expressions. EViews contains an extensive library of built-in operators and functions that allow you to perform complicated mathematical operations on your data with just a few keystrokes. In addition to supporting standard mathematical and statistical operations, EViews provides a number of specialized functions for automatically handling the leads, lags and differences that are commonly found in time series data.

An EViews expression is a combination of numbers, series names, functions, and mathematical and relational operators. In practical terms, you will use expressions to describe all mathematical operations involving EViews objects.

As in other programs, you can use these expressions to calculate a new series from existing series, to describe a sample of observations, or to describe an equation for estimation or forecasting. However, EViews goes far beyond this simple use of expressions by allowing you to use expressions virtually anywhere you would use a series. We will have more on this important feature shortly, but first, we describe the basics of using expressions.

### Operators

EViews expressions may include operators for the usual arithmetic operations. The operators for addition (+), subtraction (-), multiplication (\*), division (/) and raising to a power (^) are used in standard fashion so that

```
5 + 6 * 7.0 / 3
7 + 3e-2 / 10.2345 + 6 * 10^2 + 3e3
3^2 - 9
```

are all valid expressions. Notice that explicit numerical values may be written in integer, decimal, or scientific notation.

In the examples above, the first expression takes 5 and adds to it the product of 6 and 7.0 divided by 3 ( $5 + 14 = 19$ ); the last expression takes 3 raised to the power 2 and subtracts 9 ( $9 - 9 = 0$ ). These expressions use the standard order of evaluation, described below.

The “-” and “+” operators are also used as the unary minus (negation) and unary plus operators. It follows that

```
2-2
-2+2
2+++++++++--2
2---2
```

all yield a value of 0.

EViews follows the usual order in evaluating expressions from left to right, with operator precedence order (from highest precedence to lowest):

- unary minus (-), unary plus (+)
- exponentiation (^)
- multiplication (\*), division (/)
- addition (+), subtraction (-)
- comparison (<, >, <=, >=, =)
- and, or

The last two sets of operators are used in logical expressions. These expressions are described below.

To enforce a particular order of evaluation, you can use parentheses. As in standard mathematical analysis, terms which are enclosed in parentheses are treated as a subexpression and evaluated first, from the innermost to the outermost set of parentheses. We strongly recommend the use of parentheses when there is any possibility of ambiguity in your expression.

To take some simple examples,

- $-1^2$ , evaluates to  $(-1)^2 = 1$  since the unary minus is evaluated prior to the power operator.
- $-1 + -2 * 3 + 4$ , evaluates to  $-1 + -6 + 4 = -3$ . The unary minus is evaluated first, followed by the multiplication, and finally the addition.
- $(-1 + -2) * (3 + 4)$ , evaluates to  $-3 * 7 = -21$ . The unary minuses are evaluated first, followed by the two additions, and then the multiplication.

- $3 * ((2+3) * (7+4) + 3)$ , evaluates to  $3 * (5*11 + 3) = 3 * 58 = 174$ .

A full listing of operators is presented in Appendix A, “Mathematical Operators and Functions” on page 565.

## Series Expressions

Much of the power of EViews comes from the fact that expressions involving series operate on every observation, or element, of the series in the current sample. For example, the series expression

$$2*y + 3$$

tells EViews to multiply every sample value of Y by 2 and then to add 3. We can also perform operations that work with multiple series. For example,

$$x/y + z$$

indicates that we wish to take every observation for X and divide it by the corresponding observation on Y, and add the corresponding observation for Z.

## Series Functions

EViews contains an extensive library of built-in functions that operate on all of the elements of a series in the current sample. Some of the functions are “element functions” which return a value for each element of the series, while others are “summary functions” which return scalars, vectors or matrices, which may then be used in constructing new series or working in the matrix language (see the *Command and Programming Reference* for a discussion of scalar, vector and matrix operations).

Most function names in EViews are preceded by the @-sign. For example, @mean returns the average value of a series taken over the current sample, and @abs takes the absolute value of each observation in the current sample.

All element functions return NAs when any input value is missing or invalid, or if the result is undefined. Functions which return summary information generally exclude observations for which data in the current sample are missing. For example, the @mean function will compute the mean for those observations in the sample that are non-missing.

A full description of the functions is presented in Appendix A, “Mathematical Operators and Functions” on page 565.

## Examples

We provide a few additional examples below without additional comment:

```
@trend(1980:1)
@movav(x,4)
@mean(x)*y + @var(x)*z
1/@sqrt(2*@acos(-1))*exp(-1/2 * x^2)
d(x,4)
```

The remainder of this chapter will provide additional examples of expressions involving functions.

## Series Elements

At times, you may wish to access a particular observation for a series. EViews provides you with a special function, `@elem`, which allows you to use a specific value of a series.

`@elem` takes two arguments: the first argument is the name of the series, and the second is the date or observation identifier.

For example, suppose that you want to use the 1980:3 value of the quarterly series Y, or observation 323 of the undated series X. Then the functions:

```
@elem(y, 1980:3)
@elem(x, 323)
```

will return the values of the respective series in the respective periods.

## Logical Expressions

A logical expression is one which evaluates to “true” or “false”. EViews also allows logical expressions to take the value “missing”, but for the moment, we will ignore this point until our discussion of missing values, below.

Logical expressions may be used as part of a mathematical operation, as part of a sample statement, or as part of an if-condition in programs.

We have already seen examples of logical expressions in our discussion of samples and sample objects. For example, we saw the sample condition:

```
incm > 5000
```

which allowed us to select observations meeting the specified condition. This is an example of a logical expression—it is true for each observation on INCM that exceeds 5000; otherwise, it is false.

More generally, logical expressions are those involving the comparison operators, `<` (less than), `>` (greater than), `<=` (less than or equal to), `>=` (greater than or equal to), `=` (equal to), `<>` (not equal to).

As described above in the discussion of samples, you can use the “and” and “or” conjunction operators to build more complicated logical expressions:

```
(incm>5000 and educ>=13) or (incm>10000)
```

It is worth noting that EViews uses the number 1 to represent true and 0 to represent false. This internal representation means that you can create complicated expressions involving logical sub-expressions. For example, you can use logical expressions to recode your data:

```
0*(inc<100)+(inc>=100 and inc<200)+2*(inc>=200)
```

which yields 0 if INC < 100, 1 if INC is greater than or equal to 100 and less than 200, and 2 for INC greater than or equal to 200.

The equality comparison operator (=) requires a bit more discussion, since the equals sign is used both in assigning values and in comparing values. We will consider this issue in a bit more depth when we discuss creating and modifying series, below. For now, note that if used in an expression,

```
incm = 2000
```

evaluates to true if income is exactly 2000, and false, otherwise.

## Leads, Lags, and Differences

It is easy to work with lags or leads of your series. Simply use the series name, followed by the lag or lead enclosed in parentheses. Lags are specified as negative numbers and leads as positive numbers so that

```
income(-4)
```

is the fourth lag of the income series, while

```
sales(2)
```

is the second lead of sales.

While EViews expects lead and lag arguments to be integers, there is nothing to stop you from putting non-integer values in the parentheses. EViews will automatically convert the number to an integer; you should be warned, however, that the conversion behavior is not guaranteed to be systematic. If you must use non-integer values, you are strongly encouraged to use the `@round`, `@floor`, or `@ceil` functions to control the lag behavior.

In many places in EViews, you can specify a range of lead or lag terms. For example, when estimating equations, you can include expressions of the form

```
income(-1 to -4)
```

to represent all of the INCOME lags from 1 to 4. Similarly, the expressions

```
sales sales(-1) sales(-2) sales(-3) sales(-4)
sales(0 to -4)
sales(to -4)
```

are equivalent methods of specifying the level of SALES and all lags from 1 to 4.

EViews also has several built-in functions for working with difference data in either levels or in logs. The “D” and “DLOG” functions will automatically evaluate the differences for you. For example, instead of taking differences explicitly,

```
income - income(-1)
log(income) - log(income(-1))
```

you can use the equivalent expressions,

```
d(income)
dlog(income)
```

You can take higher order differences by specifying the difference order. For example, the expressions

```
d(income,4)
dlog(income,4)
```

represent the fourth-order differences of INCOME and log(INCOME).

If you wish to take seasonal differences, you should specify both the ordinary, and a seasonal difference term:

```
d(income,1,4)
dlog(income,1,4)
```

are first order differences with a seasonal difference at lag 4. If you want only the seasonal difference, specify the ordinary difference term to be 0:

```
d(income,0,4)
dlog(income,0,4)
```

Mathematical details are provided in Appendix A, “Mathematical Operators and Functions” on page 565.

## Missing Data

Occasionally, you will encounter data that are not available for some periods or observations, or you may attempt to perform mathematical operations where the results are undefined (e.g., division by zero, log of a negative number). EViews uses the code NA (not available) to represent these missing values.



For the most part, you will never have to worry about NAs. EViews will generate NAs for you when appropriate, and will automatically exclude observations with NAs from statistical calculations. For example, if you are estimating an equation, EViews will use the set of observations in the sample that have no missing values for the dependent and all of the independent variables.

There are, however, a few cases where you will need to work with NAs so you should be aware of some issues.

When you perform operations using multiple series, there may be alternative approaches for handling NAs. EViews will usually provide you with the option of *casewise exclusion* (common sample) or *listwise exclusion* (individual sample). With casewise exclusion, only those observations for which *all* of the series have non-missing data are used. This rule is always used, for example, in equation estimation. For listwise exclusion, EViews will use the maximum number of observations possible for each series, excluding observations separately for each series in the list of series. For example, when computing descriptive statistics for a group of series, you have the option to use a different sample for each series.

If you must work directly with NAs, just keep in mind that EViews NAs observe all of the rules of IEEE NaNs. This means that performing mathematical operations on NAs will generate missing values. Thus, each of the following expressions will generate missing values:

```
@log(-abs(x))
1/(x-x)
(-abs(x))^(1/3)
3*x + na
exp(x*na)
```

The equality and inequality comparison operators will work as expected with NA values treated as if they were any other value. Thus, the expressions

```
x = 5
y <> 5
x = na
y <> na
```

will all evaluate to “true” or “false” depending upon the value of each observation of the series.

In contrast, inequality comparisons using NAs are not logically defined, and will always return NA values:

```
x >= na
y < na
```

will yield NAs for all observations, irrespective of the values of X and Y. Similarly,

```
x > 5
```

will return a “true” or “false” for non-NA values of X, and NAs where X is missing.

If used in a mathematical operation, a logical expression resulting in an NA is treated as an ordinary missing value. For example, for observations where the series X contains NAs, the mathematical expression

```
5*(x>3)
```

will yield NAs. However, if the logical expression is used as part of a sample or if-statement, NA values are treated as “false”.

```
smpl 1 1000 if x>3
```

```
smpl 1 1000 if x>3 and x<>NA
```

are equivalent since the condition `x>3` implicitly tests for NA values. One consequence of this behavior is that

```
smpl 1 1000 if x<=NA
```

will result in a sample with no observations.

Previous versions of EViews followed the same IEEE rules for missing data as EViews 3 with one important exception. In previous versions multiplying any number by zero resulted in zero. In EViews 3, the value NA times zero equals NA. Thus a recommended method of recoding (replacing) NA's in the series X to a number Y in EViews 2 will no longer work:

```
x = (x<>na)*x + (x=na)*y
```

works in Version 2, but not in Version 3. A new `@nan` function has been provided for this purpose.

```
x = @nan(x,y)
```

recodes NA values of X to take the values in the series Y. See Appendix A, “Mathematical Operators and Functions” on page 565.

## Working with Series

One of the primary uses of expressions is to generate new series from existing data or to modify the values in an existing series. Used in combination with samples, expressions allow you to perform sophisticated transformations of your data, saving the results in new or existing series objects.

To create or modify a series, select **Quick/Generate Series...** or click on the **Genr** button on the workfile toolbar. EViews opens a window prompting you for additional information.

You should enter the assignment statement in the upper edit box, and the relevant sample period in the lower edit box.

The assignment statement is actually an implicit loop over observations. Beginning with the first observation in the sample, EViews will evaluate the assignment statement for each included observation.

### Basic Assignment

You can type the series name, followed by an equals sign and then an expression. For every element of the sample, EViews will evaluate the expression on the right-hand side of the equality, and assign the value to the *destination series* on the left-hand side, creating the series if necessary.

For example, if there is no series named Y,

$$y = 2 * x + 37 * z$$

will first create the Y series and fill it with NAs. Then, for every observation in the current sample, EViews will fill each element of the Y series with the value of the expression. If Y does exist, EViews will only replace Y values in the current sample with the value of the expression. All observations not in the sample will be unchanged.

One special form of assignment occurs when the right-hand side expression is a constant expression:

$$y = 3$$

$$y = 37 * 2 + 3$$

EViews will simply loop through all of the observations in the sample and assign the value of the constant.

### Using Samples

By modifying the sample of observations used in assignment, you can splice together series using multiple **Genr** commands. For example, if we enter three **Genr** commands with different samples, first,

Upper window: `y = z`

Lower window: `@all if z <= 1 and z > -1`

followed by a **Genr** with

Upper window: `y = -2 + 3 * z`

Lower window: `@all if z > 1`



and finally,

Upper window:  $y = -.9 + .1 * z$

Lower window: `@all if z <= -1`

we can generate Y as a piecewise linear function of the series Z.

Note that while it is possible to perform these types of operations using loops and IF-statements (see the *Command and Programming Reference*), we strongly urge you to use **Genr** and sample statements wherever possible since the latter approach is much more efficient.

## Dynamic Assignment

Since EViews evaluates the assignment expression for each observation in the sample, you can perform dynamic assignment by using lagged values of the destination series on the right side of the equality. For example, suppose we have an annual workfile that ranges from 1945 to 1997. Then if we enter:

Upper window:  $y = y + y(-1)$

Lower window: `1946 1997`

EViews will replace the Y series with the cumulative sum of Y. We begin with 1946, since we do not want to transform the first value in the workfile. Then for each period, EViews will take the current value of Y and add it to the lagged value of Y. The assignment is dynamic since as we successively move on to the next period, the lagged value of Y contains the cumulative sum.

Note that this procedure destroys the original data. To create a new series with the cumulative sums, you will have to perform the assignment in two steps, first making a copy of the original series, and then performing the dynamic assignment.

## Implicit Assignment

You can make an implicit assignment by putting a simple formula on the left-hand side of the equals sign. EViews will examine your expression and select, as the destination series, the first valid series name on the left-hand side of the equality. Then for every observation in the sample, EViews will assign values using the implicit relationship. For example, if you enter:

$\log(y) = x$

EViews will treat Y as the destination series, and evaluate  $y = \exp(x)$  for every observation in the sample.

The following are examples of valid assignment statements where Y is the destination series:

```

1/y = z
log(y/x)/14.14 = z
log(@inv(y)*x) = z
2+y+3*z = 4*w
d(y) = nrnd

```

In general, EViews can solve for, or *normalize*, equations that use the following on the left-hand side of the equality: +, -, \*, /, ^, log(), exp(), sqr(), d(), dlog(), @inv().

Since **Genr** is not a general equation solver, there will be situations in which EViews cannot normalize your equation. You cannot, for example, use the assignment statement

```
@tdist(y, 3) = x
```

since @tdist() is not one of the equations that EViews knows how to solve. Similarly, EViews cannot solve for equations where the destination series appears more than once on the left side of the equality. For example, EViews cannot solve the equation,

```
x + 1/x = 5
```

In both cases, EViews will display the error message “Unable to normalize equation”.

Note that the destination series can appear on both sides of the equality. For example,

```
log(x) = x
```

is a legal assignment statement. EViews will normalize the expression and perform the assignment

```
x = exp(x)
```

so that X will be assigned the exponential of the original value of X. EViews will *not* solve for the values of X satisfying the equality  $\log(x) = x$ .

## Command Window Assignment

You can create series and assign values from the command window. First, set the sample using the sample statement, then enter the assignment statement.

There are alternative forms for the assignment statement. First, if the series does not exist, you must use either the *series* or the *genr* keywords, followed by the assignment expression. The two statements

```
series y = exp(x)
genr y = exp(x)
```

are equivalent methods of generating the series Y. Once the series has been created, subsequent assignment statements do not require the keywords:

```
smpl @all
series y = exp(x)
smpl 1950 1990 if y>300
y = y/2
```

## Examples

The following examples demonstrate how to perform some common tasks involving generating series (using the command line syntax):

- Time trend (with the value 0 in 1987:1)

```
series time = @trend(1987:1)
```

- Price index (with the value 1 in 1985:6)

```
series index = price/@elem(price, 1985:6)
```

- Dummy variables

```
series high_inc = income>12000
```

- Recode using logical expressions

```
series newser = x*(y>30) + z*(y<=30)
```

## Working with Auto-series

Another important method of working with expressions is to use an expression *in place of* a series. EViews' powerful tools for expression handling allow you to substitute expressions virtually any place you would use a series—as a series object, as a group element, in equation specifications and estimation, and in models.

We term expressions that are used in place of series as *auto-series*, since the transformations in the expressions are *automatically* calculated without an explicit assignment statement.

Auto-series are most useful when you wish to see the behavior of a function of series, but do not want to keep the transformed series, or in cases where the underlying series data change frequently. Since the auto-series expressions are automatically recalculated whenever the underlying data change, they are never out-of-date.

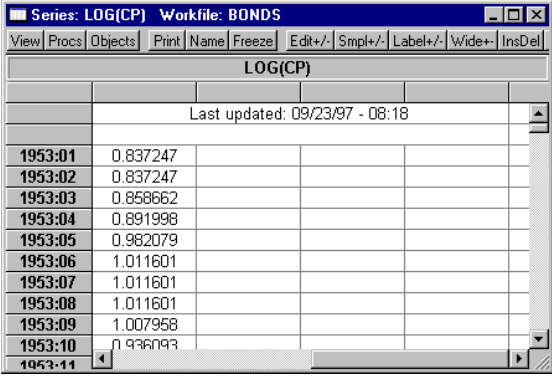
## "Creating" Auto-series

It is easy to create and use an auto-series—anywhere you might use a series name, simply enter an EViews expression. For example, suppose that you wish to plot the log of CP against time for the period 1953:01 to 1958:12. One way to plot these values is to generate a new series, say LOGCP, and then plot the values. To generate the new series, enter

```
series logcp = log(cp)
```

in the command window, open the series LOGCP and select **View/Line Graph**.

Alternatively, you can create and use an auto-series by clicking on the **Show** button, or selecting **Quick/Show...** and entering `log(cp)`. EViews will open a series window in spreadsheet view:

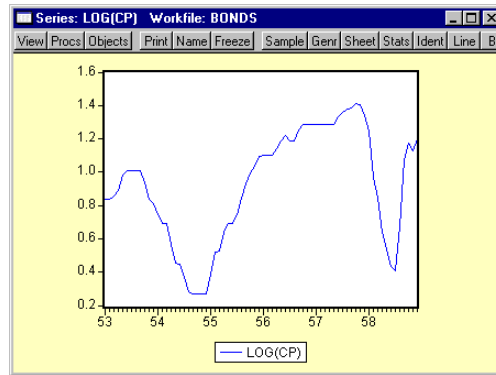


LOG(CP)	
Last updated: 09/23/97 - 08:18	
1953:01	0.837247
1953:02	0.837247
1953:03	0.858662
1953:04	0.891998
1953:05	0.982079
1953:06	1.011601
1953:07	1.011601
1953:08	1.011601
1953:09	1.007958
1953:10	0.936093
1953:11	

Note that in place of an actual series name, EViews substitutes the expression used to create the auto-series.

Furthermore, the auto-series may be treated as a standard series window so all of the series views and procedures are immediately available.

To display a time series graph of the LOG(CP) series, simply select **View/Line Graph**:

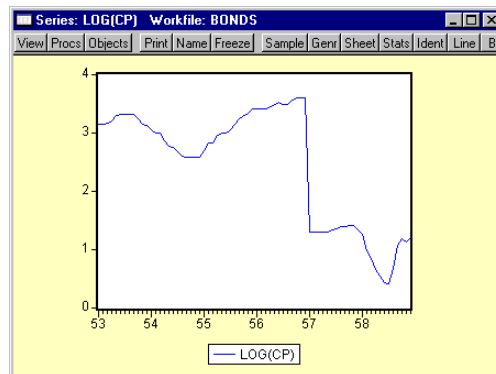


All other views and procedures are also accessible from the menus.

Note that if the data in the CP series are altered, the auto-series will reflect these changes. Suppose, for example, that we take the first four years of the CP series, and multiply it by a factor of 10:

```
smp1 1953:01 1956:12
cp = cp*10
smp1 1953:01 1958:12
```

the auto-series graph will automatically change to reflect the new data:

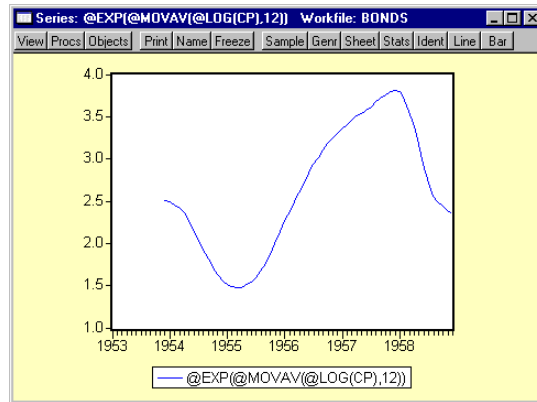


Similarly, you may use an auto-series to compute a 12 period, backward-looking, geometric moving average of the *original* CP data. The command:

```
show @exp(@movav(@log(cp), 12))
```

will display the auto-series containing the geometric moving average:



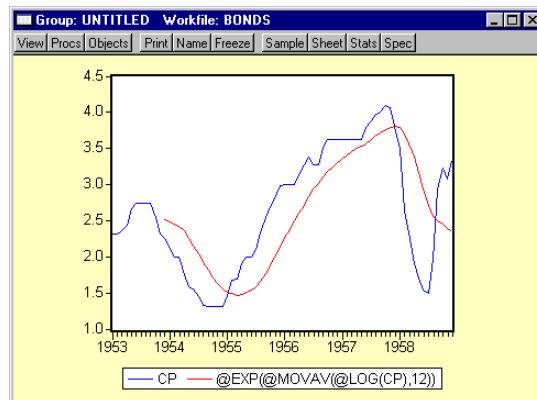


### Using Auto-series in Groups

One of the more useful ways of working with auto-series is to include them in a group. Simply create the group as usual, using an expression in place of a series name, as appropriate. For example, if you select **Objects/New Object/Group**, and enter

```
cp @exp(@movav(@log(cp),12))
```

you will create a group containing two series: the ordinary series CP, and the auto-series representing the geometric moving average. We can then use the group object graphing routines to compare the original, with the smoothed series:



The discussion in the section “Working with Groups of Series” on page 96 below describes other useful techniques for working with auto-series.

## Using Auto-Series in Estimation

One method of using auto-series in estimation is to allow expressions as right-hand side variables. Thus, you could estimate an equation with  $\log(x)$  or  $\exp(x+z)$  as an explanatory variable.

EViews goes a step beyond this use of auto-series, by allowing you to use auto-variables as the dependent variable in estimation. Thus, if you want to regress the log of  $Y$  on explanatory variables, you don't have to create a new variable  $\text{LOGY}$ . Instead, you can use the expression  $\log(y)$  as your dependent variable.

When you forecast using an equation with an auto-series dependent variable, EViews will, if possible, forecast the untransformed dependent variable and adjust the estimated confidence interval accordingly. For example, if the dependent variable is specified as  $\log(y)$ , EViews will allow you to forecast the level of  $Y$ , and will compute the asymmetric confidence interval. See Chapter 14, “Forecasting from an Equation”, beginning on page 305 for additional details.

## Working with Groups of Series

EViews provides specialized tools for working with groups of series that are held in the form of a group object. In Chapter 4 we used groups to import data from spreadsheets. Briefly, a group is a collection of one or more series identifiers or expressions. Note that a group does not contain the data in the individual series, only references to the data in the series.

To create a group, you can follow the procedure described in Chapter 8—select **Objects/New Object/Group** and fill in the dialog with names of series and auto-series. Or you can select **Show** from the workfile toolbar and fill out the dialog. Alternatively, type the command `group` in the command window, followed by a name to be given to the group and then the series and auto-series names:

```
group macrolist gdp invest cons
```

creates the group `MACROLIST` containing the series `GDP`, `INVEST` and `CONS`. Similarly,

```
group altlist log(gdp) d(invest) cons/price
```

creates the group `ALTLIST` containing the log of the series `GDP`, the first difference of the series `INVEST`, and the `CONS` series divided by the `PRICE` series.

There are a few features of groups that are worth keeping in mind:

- A group is simply a list of series identifiers. It is not a copy of the data in the series. Thus, if you change the data for one of the series in the group, you will see the changes reflected in the group.

- If you delete a series from the workfile, it will disappear from any group that originally included the series. If the deleted series is the only series in a group, the group will also be deleted.
- Renaming a series changes the reference in every group containing the series.

## Accessing Individual Elements

Groups, like other EViews objects, contain their own views and procedures. We describe groups in greater detail in Chapter 8, but for now, note that you can access the individual elements of a named group as individual series.

To refer the  $n$ -th series in the group, simply append “(n)” to the group name. For example, consider the MACROLIST group, defined above. The expression MACROLIST(1) may be used to refer to GDP and MACROLIST(2) to refer to INVEST.

You can work with MACROLIST(1) as though it were any other series in EViews. You can display the series by clicking on the **Show** button on the toolbar and entering MACROLIST(1). You can include GDP in another group directly or indirectly. A group which contains

```
macrolist(1) macrolist(2)
```

will be identical to a group containing

```
gdp invest
```

We can also use the individual group members as part of expressions in generating new series:

```
series realgdp = macrolist(1)/price  
series y = 2*log(macrolist(3))
```

or in modifying the original series:

```
series macrolist(2) = macrolist(2)/price
```

Note that in this latter example the `series` keyword is required, despite the fact that the INVEST series already exists.

Other tools allow you to retrieve the number of series in a group by appending “.@count” to the group name:

```
scalar numgroup = macrolist.@count
```

and to retrieve the names of each of the series using “.@seriesname”. These tools are described in greater detail in the *Command and Programming Reference*.

### Illustration

Access to the members of a group, when paired with auto-series, provides a powerful set of tools for working with series data. As we saw above, auto-series provide you with dynamic updating of expressions. If we use the auto-series expression:

```
log(y)
```

the result will be automatically updated whenever the contents of the series Y changes.

A potential drawback of using auto-series is that expressions may be quite lengthy. For example, the two expressions:

```
log(gdp)/price + d(invest) * (cons + invest)
12345.6789 * 3.14159 / cons^2 + dlog(gdp)
```

are not suited to use as auto-series if they are to be used repeatedly in other expressions.

You can employ group access to make this style of working with data practical. First, create groups containing the expressions:

```
group g1 log(gdp)/price+d(invest)*(cons+invest)
group g2 12345.6789*3.14159/cons^2+dlog(gdp)
```

If there are spaces in the expression, the entire contents should be enclosed in parentheses.

You can now refer to the auto-series as G1(1) and G2(1). You can go even further by combining the two auto-series into a single group:

```
group myseries g1(1) g2(1)
```

and then referring to the series as MYSERIES(1) and MYSERIES(2). If you wish to skip the intermediate step of defining the subgroups G1 and G2, make certain that there are no spaces in the subexpression or that it is enclosed in parentheses. For example, the two expressions in the group ALT SERIES,

```
group altseries (log(gdp)/price) 3.141*cons/price
```

may be referred to as ALT SERIES(1) and ALT SERIES(2).

## Working with Scalars

Scalar objects are different from series and groups in that they have no window views. Scalars are created by commands of the form

```
scalar scalar_name = number
```

where you assign a number to the scalar name. The number may be an expression or special functions that return a scalar.

To examine the contents of a scalar, you may enter the command `show`, followed by the name of the scalar. EViews will display the value of the scalar in the status line at the bottom of the EViews window. For example,

```
scalar logl1=eq1.@logl  
show logl1
```

stores the log likelihood value of the equation object named EQ1 in a scalar named LOGL1, and displays the value in the status line. Alternatively, double clicking on the scalar name in the workfile window displays the value in the status line.



## Chapter 6. EViews Databases

---

An EViews database resembles a workfile in that it is used to contain a collection of EViews objects. It differs from a workfile in two major ways. First, unlike a workfile, the entire database need not be loaded into memory in order to access an object inside it; an object can be fetched or stored directly to the database on disk. Second, the objects in an EViews database are not restricted to being of a single frequency or range. A database could contain a collection of annual, monthly, and daily series, all with different numbers of observations.

EViews databases also differ from workfiles in that they support powerful query features which can be used to search through the database to find a particular series or a set of series with a common property. This makes databases ideal for managing large quantities of data.

EViews also supports *database links*. A database link behaves similarly to a database, but instead of containing the data itself, it provides a link to data on a remote server. EViews can create database links to Standard & Poor's / DRI data services, giving you access to hundreds of thousands of historical time series directly from within EViews.

### An Overview

An EViews database is a set of files containing a collection of EViews objects. In this chapter we describe how to:

- Create a new database.
- Work with objects in the database, including how to store and fetch objects into workfiles, and how to copy, rename and delete objects in the database.
- Use auto-series to work with data directly from the database without creating a copy of the data in the workfile.
- Use the database registry to create shortcuts for long database names and to set up an search path for series names not found in the workfile.
- Perform a query on the database to get a list of objects with particular properties.
- Maintain a database with operations such as packing, copying, and repairing.
- Work with remote database links to access data from remote sites.

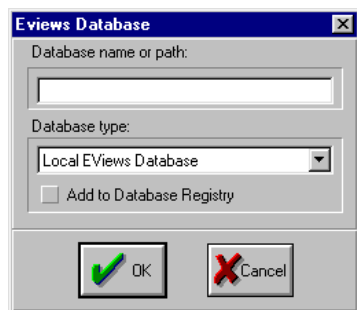
## Database Basics

### What is an EViews Database?

An EViews database consists of a set of files on disk. There is a main file with the extension .EDB which contains the actual object data, and a number of index files with extensions such as .E0, .E1A and .E1B which are used to speed up searching operations on the database. In normal use, EViews manages these files for the user, so there is no need to be aware of this structure. However, if you are copying, moving, renaming, or deleting an EViews database from outside of EViews (using Windows Explorer for example), you should perform the operation on both the main database file and all the index files associated with the database. (If you accidentally delete or damage an index file, EViews can regenerate it for you from the main data file using the repair command (see “Maintaining the Database” on page 127, below).

The fact that EViews databases are kept on disk rather than in memory has some important consequences. Any changes made to a database cause immediate changes to be made to the disk files associated with the database. Therefore, unlike workfiles, once a change is made to a database, there is no possibility of discarding the change and going back to the previously saved version. Because of this, you should take care when modifying a database, and should consider keeping regular backup copies of databases which you modify frequently.

### Creating a Database



To create a database, simply select **File/New/Database...** from the main menu. .

To create a local EViews database, simply enter a name for the database in the field labeled **Database name or path**, then click on the button marked **OK**. This will create a new database in the current EViews path.

If you would like to create a database in another directory, simply add a pathname before the database name. This can be either a relative pathname (such as `new\test`), in which case EViews will try to create the database in the

subdirectory of the current path, or an absolute pathname (such as `f:\public\test`), in which case EViews will try to create the database in the specified absolute path.

The **Database Type** field allows you to create different types of databases; see “Connecting to Remote Databases” on page 130 for details on creating a link to a remote Standard & Poor’s/DRI database.

The **Add to Database Registry** provides a convenient way to register a new database in the database registry; see “The Database Registry” on page 118 below for details.



A database can also be created from the command line or in a program using the command:

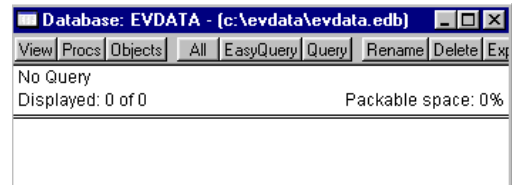
```
dbcreate db_name
```

where db\_name is the name of the database using the same rules given above.

## The Database Window

When you create a new database, a database window will open on the screen.

The database window provides a graphical interface which allows you to query the database, copy-and-paste objects to and from your workfile, and perform basic maintenance on the database. Note that many database operations can also be carried out directly without first opening the database window.



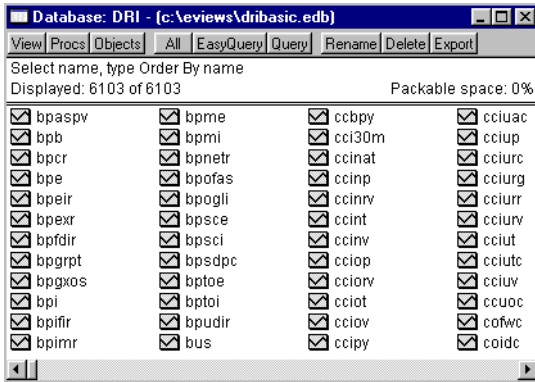
To open a database window for an existing database, select **File/Open/Database...** from the main menu. The Open dialog appears with the filename box containing all files in the selected directory with extension .EDB. Simply select the database you wish to open, and click on the button marked **Open**. A new window should appear representing the open database.

From the command line or in a program, you can open a database window by typing:

```
dbopen db_name
```

Unlike a workfile window, a database window does not display the contents of the database when it is first opened, although it does tell you how many objects are in the database. The second line of the window text shows the number of objects currently displayed (zero when the window is first opened) followed by the total number of objects stored in the database.

You can bring up an alphabetical listing of every object in the workfile by clicking on the **All** button:



As for a workfile, each object is preceded by a small icon that identifies the type of the object. When performing an **All** query, no other information about the object is visible. However, by double-clicking on an object you can bring up a full description of the object including its name, type, modification date, frequency, start and end date (for series), and label.

For large databases, the **All** button generally displays too many objects and not enough information about each object. The database query features described below allow you to control pre-

cisely which objects should be displayed, and what information about each object should be visible. The text form of the query currently being displayed is always visible in the top line of the database window.

The “Packable space” field in the database window displays the percentage of unused space in the database that can be recovered by a database pack operation; see “Packing the Database” below for details.

A brief technical note: having a database window open in EViews does not keep a file open at the operating system level. EViews only opens files when it is working with those files. Consequently, multiple users can have a database open at the same time and can often perform operations simultaneously. There are some limits imposed by the fact that one user cannot read from a database that another user is writing to at the same time. However, EViews will detect this situation and continue to retry the operation until the database becomes free. If the database does not become available within a specified time, EViews will generate an error stating that a ‘sharing violation’ on the database has occurred.

## The Default Database

In order to simplify common operations, EViews uses the concept of a *default database*. The default database is used in several places, the most important of which is as the default source or destination for store or fetch operations when an alternative database is not explicitly specified.

The default database is set by opening a new database window, or by clicking on an already open database window if there are multiple databases open on the screen. The name of the default database is listed in the status line at the bottom of the main EViews window (see Chapter 1 for details). The concept is similar to that of the current workfile with one exception: when there are no currently open databases there is still a default database; when there are no currently open workfiles, the current workfile is listed as “none.”

## EViews .DB? files

The database features described in this chapter are new to Version 3 of EViews. Previous versions of EViews and MicroTSP supported a much more limited set of database operations. Objects could be stored on disk in individual files, with one object per file. Essentially, the disk directory system was used as a database and each database entry had its own file. These files had the extension .DB for series, and .DB followed by an additional character for other types of objects. EViews refers to these collectively as .DB? files.

While the new database features added to EViews provide a superior method of archiving and managing your data, we have continued to support .DB? files to provide backward compatibility, and as a convenient method of distributing data to other programs. Series .DB files are now supported by a large number of programs including TSP, RATS, and SHAZAM. Also, some organizations such as the National Bureau of Economic Research (NBER), now distribute data in .DB format.

## Working with Objects in Databases

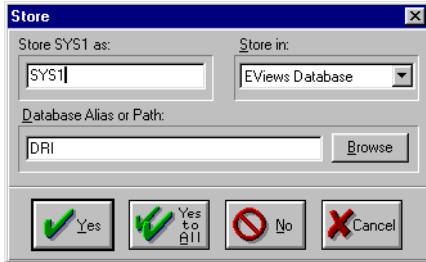
Since databases are simply containers of other EViews objects, most of your work with databases will involve moving objects into and out of them. The sections on storing, fetching and exporting objects discuss different ways of doing this.

You will also need to manage the objects inside a database. You can create duplicate copies of objects, change their names, or remove them from the database entirely. The sections on copying, renaming and deleting discuss how these operations can be carried out.

Moving a series between a database and a workfile often requires converting the frequency of the data to match that of the workfile. The rules by which conversion is performed are discussed in the section on frequency conversion.

### Storing Objects in the Database

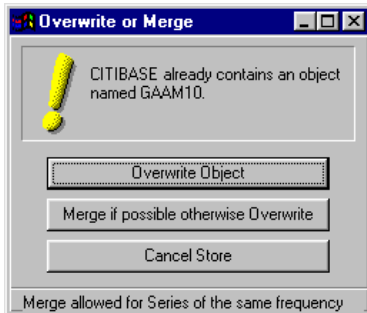
An object can be stored in a database in a number of ways. If you have a workfile open on the screen and would like to store objects contained inside it into a database, just select the objects from the workfile window with the mouse, then click on the **Store** button in the workfile toolbar. A sequence of dialogs will come up, one for each object selected, which provide a number of options for renaming the object and determining where the object should be stored.



By default, the object will be stored in the current default database with the same name as in the workfile. Click **Yes** to store this object, or **Yes-to-All** to store all the selected objects using these settings.

If you would like to store the object with a different name, simply type the new name over the old name in the edit box labeled **Store object\_name as:** (where object\_name is the name of the object currently being stored). If you would like to store the object in a different database, either type in the

name of the new database in the text box marked **Database Alias or Path:** (see “The Database Registry” on page 118 below for an explanation of database aliases), or click on the button marked **Browse** to select the database name interactively. To store the object to disk as an EViews .DB? file, click on the arrow to the right of the field labeled **Store in:** and select **Individual .DB? files**. You may then specify a path in which to place the file using the field labeled **Path** for DB files.



If there is already an existing object in the database with the same name, EViews will display a dialog. The first and last of the three options are self explanatory. The second option can only be used if the object you are storing from the workfile and the object already in the database are both series of the same frequency. In this case, EViews will merge the data from the two series so that the new series in the database has all the observations from the series being stored, as well as any observations from the existing series which lie outside this range. For example, if the existing series in the database

is an annual series from 1950 to 1990, and the series being stored is an annual series from 1980 to 1995, the new series will run from 1950 to 1995, with data from the existing series for 1950 to 1979, and data from the new series for 1980 to 1995.

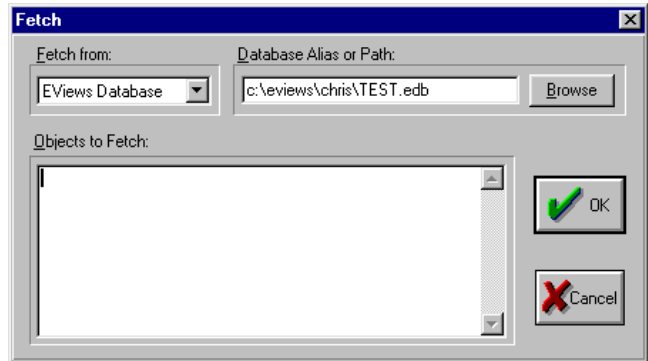
## Fetching Objects from the Database

There are a number of ways to fetch objects from a database, most of which are similar to the methods for storing.

The first method is to click on the button marked **Fetch** on the toolbar of the workfile into which you would like to fetch the object. A dialog will come up which is similar to the dialog for store:

The dialog allows you to specify the names of the objects to fetch, and the database or directory from which to retrieve them.

Enter the names of the objects you would like to fetch in the field **Objects to Fetch**. Alternatively, you can use the mouse to select objects from the workfile window before



clicking on the **Fetch** button, in which case the names of these objects will appear automatically. The fields labeled **Database Alias or Path** and **Fetch from** are the same as for the store dialog with one exception. In addition to **EViews Database** and **Individual .DB? files**, **Fetch from** has an option titled **Search Databases**. This option tells EViews to search multiple databases for objects which match the specified names. To use this option you must first define a search order in the database registry; see “The Database Registry” on page 118 below.

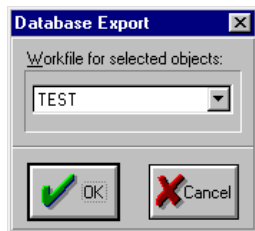
When you click on **OK**, EViews will fetch all the objects. If an object which is being fetched is already contained in the workfile, a dialog will appear asking whether to replace the object or not. Click on **Yes** to replace the object in the workfile or **No** to leave the object in the workfile unchanged.

Because a workfile has a fixed frequency and range, fetching a series into a workfile may cause the data in the series to be modified to match the frequency and range of the workfile (see “Frequency Conversion” on page 111 below for details). Because of this, loading a series into a workfile then saving it back into the database can cause truncation and frequency conversion of the series stored in the database. You should be careful that this does not happen accidentally.

**Objects/Update selected from DB...** from the workfile toolbar is the same as **Fetch** except that there is no overwrite warning message. If the object in the database is the same type as the one in the workfile, it is automatically overwritten. If it is of a different type, the fetch does not proceed. Update is also available from the **Objects** button in individual object windows.

## Database Export

You can also move data into a workfile from the database window. From an open database window, select the objects you would like to copy using the mouse, then click on the button marked **Export** in the toolbar at the top of the database window. A dialog titled Database Export will appear on the screen:



When you click on the down arrow on the right of the field labeled **Workfile for selected objects:**, a list of all workfiles that are currently open should appear. Choose the workfile into which you would like to copy the objects. Clicking on the button marked **OK** will copy the selected objects to the chosen workfile.

In the list of open workfiles is an extra option called **New Workfile**. Choosing this option allows you to create a new workfile containing the objects you have selected. After you click on **OK** a dialog will appear in which you can set the frequency and range of the workfile to be created. The default frequency is set to the lowest frequency of any of the objects selected, and the default range is set to cover all the data points contained in the objects. Clicking on **OK** will open a new workfile window and copy the selected objects into it, performing frequency conversions where necessary.

## Copying Objects

In addition to the above methods for moving objects, EViews provides general support for the copying and pasting of objects between any two EViews container objects (workfiles or databases). You can use these features to move objects between two databases or between two workfiles, to create duplicate copies of objects within a workfile or database, or as an alternative method for store and fetch.

For copying objects between containers, the procedure is very similar no matter what types of container objects are involved. Before you start, make sure that the windows for both containers are open on the screen. In the container *from which* you would like to copy the objects, select the objects then click on **Edit/Copy** in the EViews program menu. Click on the container object *into which* you would like to paste the objects, then select **Edit/Paste** from the EViews program menu.

Depending on the types of the two containers, you will be presented with one or more dialogs. If you are performing a fetch or store operation, the same dialogs will appear as if you had carried out the operations using the toolbar buttons on the workfile window.

You can perform similar operations, as well as creating duplicate copies of an object within a container, by using the object copy procedure. From the main menu select **Object/Copy** (this may appear as **Object/Copy selected...**). The Object Copy dialog will be displayed.

The **Source** field specifies the object or objects you would like to copy, the **Destination** field specifies where you would like to copy them and what names they should be given.

The **Source** field should be filled in with an expression of the form:

```
source_db::source_pattern
```

where `source_db::` is optional, and indicates which database the objects should be copied from (if no database name is supplied, the source is taken to be the default workfile), and `source_pattern` is either a simple object name, or a pattern which may include the wildcard characters `?` which matches any single character, and `*` which matches zero or more characters (see “Copy and Rename with Wildcard Characters” on page 110 below).

The **Destination** field should be filled in with an expression of the form:

```
dest_db::dest_name
```

where `dest_db::` is again optional, and indicates which database the objects should be copied to (if no database name is supplied, the destination is taken to be the default workfile), and `dest_name`, which is also optional, is the name to be given to the new copy of the object. If no name is given, the object will be copied with its existing name.

For example, to copy an object from the database DB1 to the database DB2, keeping the existing name, you would fill in the dialog:

```
source:      db1::object_name
destination: db2::
```

where `OBJECT_NAME` is the original name as displayed by EViews.

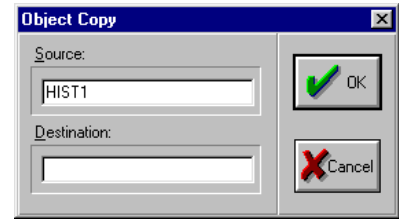
To copy all the objects in the database DB1 beginning with the letter X into the current workfile, you would fill in the dialog

```
source:      db1::x*
destination:
```

To make a duplicate copy of the object named ABC in the database DB1, giving it the new name XYZ you would fill in the dialog:

```
source:      db1::abc
destination: db1::xyz
```

EViews supports the use of wildcard characters (`?` for a single character match and `*` for a pattern match) in destination specifications when using the `copy` and `rename` commands. Using



this feature, you can copy or rename a set of objects whose names share a common pattern in a single operation.

## Renaming Objects in the Database

You may rename an object in the database by selecting the object in an open database window, then clicking on the button marked **Rename** in the database window toolbar. A dialog will come up in which you can modify the existing name or type in a new name. Only one object can be renamed at a time. There is also a `rename` command. See the *Command and Programming Reference* for details.

## Copy and Rename with Wildcard Characters

EViews supports the use of wildcard characters (? for a single character match and \* for a pattern match) in destination specifications when using the `copy` and `rename` commands. Using this feature, you can copy or rename a set of objects whose names share a common pattern in a single operation. This can be useful for managing series produced by model simulations, series corresponding to pool cross-sections, and any other situation where you have a set of objects which share a common naming convention.

A destination wildcard pattern can be used only when a wildcard pattern has been provided for the source, and the destination pattern must always conform to the source pattern in that the number and order of wildcard characters must be exactly the same between the two. For example, the following patterns

x*	y*
*c	b*
x*12?	yz*f?abc

conform to each other, while these patterns do not

a*	b
*x	?y
x*y*	*x*y*

When using wildcards, the new destination name is formed by replacing each wildcard in the destination pattern by the characters from the source name that matched the corresponding wildcard in the source pattern. Some examples should make this principle clear:



Source Pattern	Destination Pattern	Source Name	Destination Name
*_base	*_jan	x_base	x_jan
us_*	*	us_gdp	gdp
x?	x?f	x1	x1f
*_*	**f	us_gdp	usgdpf
??*f	??_*	usgdpf	us_gdp

Note, as shown in the second example, that a simple asterisk for the destination pattern does not mean to use the unaltered source name as the destination name. To copy objects between containers preserving the existing name, either repeat the source pattern as the destination pattern

```
copy x* db1::x*
```

or omit the destination pattern entirely

```
copy x* db1::
```

If you use wildcard characters in the source name and give a destination name without a wildcard character, EViews will keep overwriting all objects which match the source pattern to the name given as destination.

## Deleting Objects from the Database

To delete an object in the database, select the objects in an open database window, then click on the button marked **Delete** on the database window toolbar. There is also a `delete` command which allows pattern matching of names. See the *Command and Programming Reference* for details.

## Frequency Conversion

As already mentioned, a database can contain objects with many different frequencies. To work with data, however, you will generally need to construct a workfile in which all data share a common frequency. When you fetch or copy a series to the workfile which does not match the frequency of the workfile, EViews automatically converts the data to the workfile frequency.

There are two main types of frequency conversion. If the series has a lower frequency than the workfile (the series is less frequently observed than data in the workfile), EViews provides several methods for performing frequency conversion.

### High Frequency To Low Frequency Conversion

If the series has a higher frequency than the workfile, you may choose between a number of different conversion methods. There are six main alternatives:

- **Average observations**
- **Sum observations**
- **First observation**
- **Last observation**
- **Maximum observation**
- **Minimum observation**

In addition, there is an option **Conversion propagates NAs** which determines how missing data should be handled when carrying out the calculations. If propagate NAs is checked and a missing value appears anywhere in a calculation, the result for that period will be an NA. If propagate NAs is not checked, the calculations will be performed ignoring the missing values (although if all values for the period are missing, the result for that period will still be an NA).

For example, if you choose the **First observation** conversion method and fetch a daily series into a quarterly workfile, the value of the first quarter will be taken from the January 1 observation of the corresponding year. If no data exists for January 1, and NAs propagate is selected, an NA will be returned for the entire quarter. If NAs propagate is not selected, the quarter will be assigned the first available observation between January 1 and March 31.

If you fetch or copy an undated series into a dated workfile, the data will be copied into the workfile sequentially without any conversion, beginning at the starting observation number of the undated series (generally the first observation).

You can adjust the conversion method in three ways: as an option to the command line, as a series setting, and as a global option. The settings are applied in the following order:

- If you are using the `copy` or `fetch` command and you provide an option to set the conversion method (see the *Command and Programming Reference* for details), then EViews will use this method for all of the series listed in the command.
- Otherwise, if a series has a particular frequency conversion option set, (choose **View/Conversion Options...** from the series object menu to view or change the option) that conversion method will be used for that particular series.
- Otherwise, the global option setting will be used (choose **Options/Frequency Conversion - Dates...** from the main menu to view or change the option). The global option is initially set to average the observations and to not propagate NAs.

As an example of controlling frequency conversion, say you have daily data consisting of high, low, and close series for a particular stock, from which you would like to construct a monthly workfile. If you use the default frequency conversion methods, the monthly workfile will contain series which are an average of the daily observations. This is unlikely to be what you want. By

setting the frequency conversion method of the high series to **Maximum observation**, of the low series to **Minimum observation**, and of the close series to **Last observation**, you can create a monthly workfile which will correctly follow the definitions of the series.

### Low Frequency to High Frequency

EViews also provides a number of different interpolation methods for dealing with the case where the series being brought into the workfile has a lower frequency than the workfile. Since observing a series at a lower frequency provides fundamentally less information than observing the same series at a higher frequency, it is in general impossible to perfectly recover the high frequency series from the low frequency data. Consequently, the results from EViews' interpolation methods should be considered as suggestive rather than as providing the true values of the underlying series

EViews supports the following interpolation methods:

- **Constant:** Constant with sum or average matched to the source data.
- **Quadratic:** Local quadratic with sum or average matched to the source data.
- **Linear:** Linear with last observation matched to the source data.
- **Cubic:** Cubic spline with last observation matched to the source data.
- **No conversion:** Disallow conversion (generates error message if mixed frequency).

Using an interpolation method which matches the average means that the average of the interpolated points for each period is equal to the source data point for that period. Similarly if the sum is matched, the interpolated points will sum to the source data point for the period, and if the last observation is matched, the last interpolated point will equal the source data point for the period.

For all methods, all relevant data from the low frequency series is used when forming the high frequency series, even if the range of the destination covers only part of the source.

The following describes the different methods in greater detail.

- **Constant: match average, Constant: match sum** — These two methods assign the same value to all observations in the high frequency series associated with a particular low frequency period. In one case, the value is chosen so that the average of the high frequency observation matches the low frequency observation (the value is simply repeated), in the other case, the value is chosen so that the sum of the high frequency observations matches the low frequency observation (the value is divided by the number of observations).
- **Quadratic: match average, Quadratic: match sum** — These two methods fit a local quadratic polynomial for each observation of the low frequency series, then use this polynomial to fill in all observations of the high frequency series associated with the period. The qua-

dratic polynomial is formed by taking sets of three adjacent points from the source series and fitting a quadratic so that either the average or the sum of the high frequency points match to the low frequency data actually observed. For most points, one point before and one point after the period currently being interpolated are used to provide the three points. For end points, the two periods are both taken from the one side where data is available.

This method is a purely local method. The resulting interpolation curves are not constrained to be continuous at the boundaries between adjacent periods. Because of this, the method is better suited to situations where relatively few data points are being interpolated and the source data is fairly smooth.

- **Linear: match last** — This method assigns each value in the low frequency series to the last high frequency observation associated with the low frequency period, then places all intermediate points on straight lines connecting these points.
- **Cubic: match last** — This method assigns each value in the low frequency series to the last high frequency observation associated with the low frequency period, then places all intermediate points on a natural cubic spline connecting all the points.

A natural cubic spline is defined by the following properties:

1. Each segment of the curve is represented by a cubic polynomial.
2. Adjacent segments of the curve have the same level, first derivative and second derivative at the point where they meet.
3. The second derivative of the curve at the two global end points is equal to zero (this is the 'natural' spline condition).

Cubic spline interpolation is a global interpolation method so that changing any one point (or adding an additional point) to the source series will affect all points in the interpolated series.

## Store, Fetch, and Copy of Group Objects

A group object in EViews is essentially a list of series names that form the group. The data of each series are contained in the series object, not in the group object. When you do a store, fetch, or copy operation on a group object, an issue arises as to whether you want to do the operation on each of the series or to the group definition list.

### Storing a Group Object

When you store a group object to a database, there are four available options:

- Store the group definition and the series as separate objects: stores the group object (only its definition information) and each of its series as separate objects in the database. If any of the series already exist in the database, EViews will ask whether to overwrite the existing series if in interactive mode, and will error if in batch mode.

- Store the group definition and the series as one object: stores each series within the group object. A group object that contains series data will have an icon G + in the database directory window. A group object with only its definition information will have the usual icon G. If you use this option, you can store two different series with the same name (with one of the series as member of a group).
- Store only the series (as separate objects): only stores each series as separate objects in the database. If you want to store a long list of series into a database, you can create a temporary group object that contains those series and issue the store command only once.
- Store only the group definition: stores only the group definition information; none of the series data are stored in the database. This option is useful if you want to update the member data from the database but want to keep the group information (e.g. the dated data table settings) in the group.

By default, EViews will display a dialog asking you to select a group store option every time you store a group object. You can, however, instruct EViews to suppress the dialog and use the global option setting. Simply click on **Options/Database Default Storage Options...** to bring up a dialog that allows you both to set the global storage options, and to suppress the group store option dialog.

### Fetching a Group Object

When you fetch a group object to a database, there are three options available:

- Fetch both group definition and the actual series: fetches both group definition and its series as separate objects. If any of the series defined in the group is not found in the database, the corresponding series will be created in the workfile filled with NAs. If any of the series already exist in the workfile, EViews will ask whether to overwrite the existing series or not if in interactive mode, and will error if in batch mode.
- Fetch only the series in the group: only fetches each series defined in the group. If the series exists both within the group object (with a G + icon) and as a separate series object in the database, the series within the group object will be fetched.
- Fetch only the group definition: fetches only the group definition (but not the series data). If any of the series defined in the group does not exist in the workfile, EViews will create the corresponding series filled with NAs.

You can click on **Options/Database Default Storage Options...** to bring up a dialog that allows you both to set the global fetch options, and to suppress the fetch option dialog.

### Copying Group Objects between Workfiles and Databases

You can also copy groups between different containers. The options that are available will differ depending on the type of source and destination container:

- Copy from workfile to database: the same options as the store operation are available.
- Copy from database to workfile: the same options as the fetch operation are available.
- Copy from workfile to workfile: both the group definition and its series will be copied.
- Copy from database to database: if the group object contains only the group definition (with a G icon), only the group definition will be copied. If the group object also contains its series data (with a G + icon), then the group as it contains the series data will be copied and the copied group will also appear with a G + icon.

## Database Auto-Series

We have described how to fetch series into a workfile. There is an alternative way of working with databases which allows you to make direct use of the series contained in a database without first copying the series. The advantage of this approach is that you need not go through the process of importing the data every time the database is revised. This approach follows the model of auto-series in EViews as described in Chapter 5.

There are many places in EViews where you can use a series expression, such as  $\log(X)$ , instead of a simple series name and EViews will automatically create a temporary auto-series for use in the procedure. This functionality has been extended so that you can now directly refer to a series in a database using the syntax:

```
db_name::object_name
```

If you omit the database name and simply prefix the object name with a double colon like this:

```
::object_name
```

EViews will look for the object in the default database.

A simple example is to generate a new series:

```
series lgdp = log(macro_db::gdp)
```

EViews will fetch the series named GDP from the database named MACRO\_DB, and put the log of GDP in a new series named LGDP in the workfile. It then deletes the series GDP from memory, unless it is in use by another object. Note that the generated series LGDP only contains data for observations within the current workfile sample.

You can also use auto-series in a regression. For example:

```
ls log(db1::y) c log(db2::x)
```

This will fetch the series named Y and X from the databases named DB1 and DB2, perform any necessary frequency conversions and end point truncation so that they are suitable for use in the

current workfile, take the log of each of the series, then run the requested regression. Y and X are then deleted from memory unless they are otherwise in use.

The auto-series feature can be further extended to include automatic searching of databases according to rules set in the database registry (see “The Database Registry”, page 118 for details). Using the database registry you can specify a list of databases to search whenever a series you request cannot be found in the workfile. With this feature enabled, the `series` command

```
series lgdp = log(gdp)
```

looks in the workfile for a series named GDP. If it is not found, EViews will search through the list of databases one by one until a series called GDP is found. When found, the series will be fetched into EViews so that the expression can be evaluated. Similarly, the regression

```
ls log(y) c log(x)
```

will fetch Y and/or X if they are not found in the workfile. Note that the regression output will label all variables with the database name from which they were imported.

In general, using auto-series directly from the database has the advantage that the data will be completely up to date. If the series in the database are revised, you do not need to repeat the step of importing the data into the workfile. You can simply reestimate the equation or model and EViews will automatically retrieve new copies of any data which are required.

There is one complication to this discussion which results from the rules which regulate the updating and deletion of auto-series in general. If there is an existing copy of an auto-series already in use in EViews, then a second use of the same expression will not cause the expression to be reevaluated (in this case reloaded from the database); it will simply make use of the existing copy. If the data in the database have changed since the last time the auto-series was loaded, the new expression will use the old data.

One implication of this behavior is that a copy of a series from a database can persist for any length of time if it is stored as a member in a group. For example, if you type:

```
show db1::y db2::x
```

this will create an untitled group in the workfile containing the expressions `db1::y` and `db2::x`. If the group window is left open and the data in the database are modified (for example by a `store` or a `copy` command), the group and its window will not update automatically. Furthermore, if the regression

```
ls log(db1::y) c log(db2::x)
```

is run again, this will use the copies of the series contained in the untitled group; it will not refetch the series from the database..

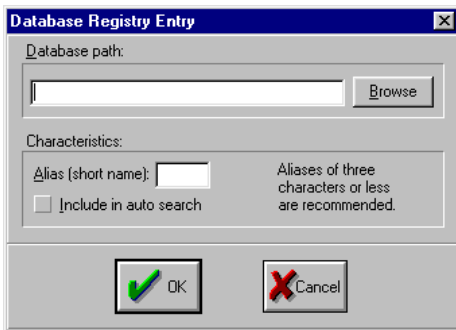
## The Database Registry

The database registry is a file on disk that manages a variety of options which control database operations. It gives you the ability to assign short alias names that can be used in place of complete database paths, and it allows you to configure the automatic searching features of EViews.

**Options/Database Registry...** from the main menu brings up the Database Registry dialog allowing you to view and edit the database registry:



The box labeled **Registry Entries** lists the databases that have been registered with EViews. The first time you bring up the dialog, the box should be empty. If you click on the **Add new entry** button, a second dialog appears.



There are three things you must specify in the dialog: the full name (including path) of the database, the alias which you would like to associate with the database, and the option for whether you wish to include the database in automatic searches.

The full name and path of the database should be entered in the top edit field. Alternatively, click the **Browse** button to select your database interactively.

The next piece of information you must provide is a *database alias*: a short name that you can use in place of the full database path in EViews commands. The database alias will also be used by EViews to label database auto-series. For example, suppose you have a database named DRIBASIC in the subdirectory DATA on drive C. The following regression command is legal but awkward:

```
ls c:\data\dribasic::gdp c c:\data\dribasic::gdp(-1)
```



Long database names such as these may also cause output labels to truncate, making it difficult to see which series were used in a procedure.

By assigning DRIBASIC the alias DRI, we can use the more readable command:

```
ls dri::gdp c dri::gdp(-1)
```

and the regression output will be labeled with the shorter names. To minimize the possibility of truncation, we recommend the use of short alias names if you intend to make use of database auto-series.

Finally you should tell EViews if you want to include the database in automatic database searches by checking the **Include in auto search** checkbox. Clicking on **OK** will return you to the main Database Registry dialog with the new entry included in the list.

Any registry entry may be edited, deleted, switched on or off for searching, or moved to the top of the search order by highlighting the entry in the list and clicking the appropriate button to the right of the list box.

The remainder of the Database Registry dialog allows you to set options for automatic database searching. The **Auto-search:** checkbox is used to control EViews behavior when you enter a command involving a series name which cannot be found in the current workfile. If this checkbox is selected, EViews will automatically search all databases that are registered for searching, before returning an error. If a series with the unrecognized name is found in any of the databases, EViews will create a database auto-series and continue with the procedure.

The last section of the dialog, **Default Database in Search Order**, lets you specify how the default database is treated in automatic database searches. Normally, when performing an automatic search, EViews will search through the databases contained in the **Registry Entries** window in the order that they are listed (provided that the **Include in auto search** box for that entry has been checked). These options allow you to assign a special role to the default database when performing a search.

- **Include at start of search order**—means that the current default database will be searched first, before searching the listed databases.
- **Include at end of search order**—means that the current default database will be searched last, after searching the listed databases.
- **Do not include in search**—means that the current default database will not be searched unless it is already one of the listed databases.

## Querying the Database

A great deal of the power of the database comes from its extensive query capabilities. These capabilities make it easy to locate a particular object, or to perform operations on a set of objects which share similar properties.

The query capabilities of the database can only be used interactively from the database window. There are two ways of performing a query on the database: the easy mode and the advanced mode. Both methods are really just different ways of building up a text query to the database. The easy mode provides a simpler interface for performing the most common types of queries. The advanced mode offers more flexibility at the cost of increased complexity.

### Easy Queries

To perform an easy query, first open the database, then click on the **EasyQuery** button in the toolbar at the top of the database window. The Easy Query dialog will appear containing two text fields and a number of check boxes:

There are two main sections to this dialog: **Select** and **Where**. The **Select** section determines which fields to display for each object that meets the query condition. The **Where** section allows you to specify conditions that must be met for an object to be returned from the query. An Easy Query allows you to set conditions on the object name, object description, and/or object type.

The two edit fields (**name** and **description**) and the set of check boxes (object **type**) in the **Where** section provide three filters of objects that are returned from the query to the database. The filters are applied in sequence (using a logical 'and' operation) so that objects in the database must meet all of the criteria selected in order to appear in the results window of the query.

The **name** and **description** fields are each used to specify a *pattern expression* that the object must meet in order to satisfy the query. The simplest possible pattern expression consists of a single pattern. A pattern can be either a simple word consisting of alphanumeric characters, or a pattern made up of a combination of alphanumeric characters and the wildcard symbols ? and \*, where ? means to match any one character and \* means to match zero or more characters. For example:

pr?d\*ction

would successfully match the words production, prediction, and predilection. Frequently used patterns include `s*` for words beginning in S, `*s` for words ending in S, and `*s*` for words containing S. Upper or lower case is not significant when searching for matches.

Matching is done on a word by word basis where at least one word in the text must match the pattern for it to match overall. Since object names in a database consist of only a single word, pattern matching for names consists of simply matching this word.

For descriptions, words are constructed as follows: each word consists of a set of consecutive alphanumeric characters, underlines, dollar signs, or apostrophes. However, the following list words are explicitly ignored: a, an, and, any, are, as, be, between, by, for, from, if, in, is, it, not, must, of, on, or, should, that, the, then, this, to, with, when, where, while. (This is done for reasons of efficiency, and to minimize false matches to patterns from uninteresting words). The three words and, or, and not are used for logical expressions; see below.

For example,

```
bal. of p'ment: seas.adj. by X11
```

is broken into the following words: bal, p'ment, seas, adj, and x11. The words of and by are ignored.

A pattern expression can also consist of one or more patterns joined together with the logical operators and, or and not in a manner similar to that used in evaluating logical expressions in EViews. That is, the keyword and requires that both the surrounding conditions be met, the keyword or requires that either of the surrounding conditions be met, and the keyword not requires that the condition to the right of the operator is not met. For example:

```
s* and not *s
```

matches all objects which contain words which begin with, but do not end with, the letter S.

More than one operator can be used in an expression, in which case parentheses can be added to determine precedence (the order in which the operators are evaluated). Operators inside parentheses are always evaluated logically prior to operators outside parentheses. Nesting of parentheses is allowed. If there are no parentheses, the precedence of the operators is determined by the following rules: not is always applied first; and is applied second; and or is applied last. For example,

```
p* or s* and not *s
```

matches all objects which contain words beginning with P, or all objects which contain words which begin with, but do not end with, the letter S.

The third filter provided in the Easy Query dialog is the ability to filter by object type. Simply select the object types which you would like displayed, using the set of check boxes near the bottom of the dialog.

## Advanced Queries

Advanced queries allow considerably more control over both the filtering and the results which are displayed from a query. Because of this flexibility, advanced queries require some understanding of the structure of an EViews database to be used effectively.

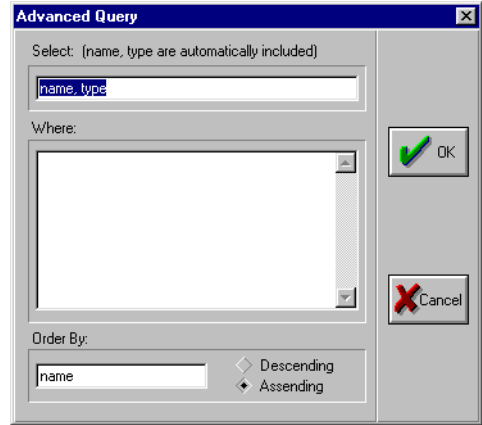
Each object in an EViews database is described by a set of fields. Each field is identified by a name. The current list of fields includes:

name	The name of the object.
type	The type of the object.
last_write	The time this object was last written to the database.
last_update	The time this object was last modified by EViews.
freq	The frequency of the data contained in the object.
start	The date of the first observation contained in the object.
end	The date of the last observation contained in the object.
obs	The number of data points stored in the series (including missing values).
description	A brief description of the object.
source	The source of the object.
units	The units of the object.
remarks	Additional remarks associated with the object.
history	Recent modifications of the object by EViews.
display_name	The EViews display name.

An advanced query allows you to examine the contents of any of these fields, and to select objects from the database by placing conditions on these fields. An advanced query can be performed by opening the database window, then clicking on the button marked **Query** in the toolbar at the top of the window. The Advanced Query dialog is displayed.

The first edit field labeled **Select:** is used to specify a list of all the fields that you would like displayed in the query results. Input into this text box consists of a series of field names separated by commas. Note that the name and type fields are always fetched automatically.

The ordering of display of the results of a query is determined by the **Order By** edit field. Any field name can be entered into this box, though some fields are likely to be more useful than others. The description field, for example, does not provide a useful ordering of the objects. The **Order By** field can be useful for grouping together objects with the same value of a particular field. For example, ordering by type is an effective way to group together the results so that objects of the same type are placed together in the database window. The **Ascending** and **Descending** buttons can be used to reverse the ordering of the objects. For example, to see objects listed from those most recently written in the database to those least recently written, one could simply sort by the field `last_write` in **Descending** order.



The **Where** edit field is the most complicated part of the query. Input consists of a logical expression built up from conditions on the fields of the database. The simplest expression is an operator applied to a single field of the database. For example, to search for all series which are of monthly or higher frequencies (where higher frequency means containing more observations per time interval), the appropriate expression is:

```
freq >= monthly
```

Field expressions can also be combined with the logical operators `and`, `or` and `not` with precedence following the same rules as those described above in the section on easy queries. For example, to query for all series of monthly or higher frequencies which begin before 1950 we could enter the expression:

```
freq >= monthly and start < 1950
```

Each field has its own rules as to the operators and constants which can be used with the field.

## Name

The name field supports the operators `<`, `<=`, `>`, `>=`, `=`, and `<>` to perform typical comparisons on the name string using alphabetical ordering. For example,

```
name >= c and name < m
```

will match all objects with names beginning with letters from C to L. The name field also supports the operator `matches`. This is the operator which is used for filtering the name field in the easy query and is documented extensively in the previous section. Note that if `matches` is used with an expression involving more than one word, the expression must be contained in quotation marks. For example,

```
name matches "x* or y*" and freq = quarterly
```

is a valid query, while

```
name matches x* or y* and freq = quarterly
```

is a syntax error because the part of the expression that is related to the `matches` operator is ambiguous.

## Type

The type field can be compared to the following object types in EViews using the `=` operator: `sample`, `equation`, `graph`, `table`, `text`, `program`, `model`, `system`, `var`, `pool`, `sspace`, `matrix`, `group`, `sym`, `matrix`, `vector`, `coef`, `series`. Relational operators are defined for the type field, although there is no particular logic to the ordering. The ordering can be used, however, to group together objects of similar types in the **Order By** field.

## Freq

The frequency field has one of the following values:

u	Undated.
a	Annual.
s	Semiannual.
q	Quarterly.
m	Monthly.
w	Weekly.
5	5 day daily.
7	7 day daily.

Any word beginning with the letter above is taken to denote that particular frequency, so that monthly can either be written `m` or `monthly`. Ordering over frequencies is defined so that a frequency with more observations per time interval is considered ‘greater’ than a series with less observations per time interval. The operators `<`, `>`, `<=`, `>=`, `=`, `<>` are all defined according to these rules. For example,

```
freq <= quarterly
```

will match objects whose frequencies are quarterly, semiannual, annual or undated.

## Start and End

Start and end dates use the following representation. A date from an annual series is written as an unadorned year number such as 1980. A date from a semiannual series is written as a year number followed by an S followed by the six month period, for example 1980S2. The same pattern is followed for quarterly and monthly data using the letters Q and M between the year and period number. Weekly, five day daily and 7 day daily data are denoted by a date in the format

```
mm/dd/yyyy
```

where m denotes a month digit, d denotes a day digit, and y denotes a year digit.

Operators on dates are defined in accordance with calendar ordering where an earlier date is less than a later date. Where a number of days are contained in a period, such as for monthly or quarterly data, an observation is ordered according to the first day of the period. For example,

```
start <= 1950
```

will include dates whose attributed day is the first of January 1950, but will not include dates which are associated with other days in 1950, such as the second, third, or fourth quarter of 1950. However, the expression

```
start < 1951
```

would include all intermediate quarters of 1950.

## Last\_write and Last\_update

As stated above, `last_write` refers to the time the object was written to disk, while `last_update` refers to the time the object was last modified inside EViews. For example, if a new series was generated in a workfile, then stored in a database at some later time, `last_write` would contain the time that the store command was executed, while `last_update` would contain the time the new series was generated. Both of these fields contain date and time information which is displayed in the format:

```
mm/dd/yyyy hh:mm
```

where m represents a month digit, d represents a day digit, y represents a year digit, h represents an hour digit and m represents a minute digit.

The operators `<`, `>`, `<=`, `>=`, `=`, and `<>` are defined on the time fields so that earlier dates and times are considered less than later dates and times. A typical comparison has the form:

```
last_write >= mm/dd/yyyy
```

A day constant always refers to twelve o'clock midnight at the beginning of that day. There is no way to specify a particular time during the day.

### Description, Source, Units, Remarks, History, Display\_name

These fields contain the label information associated with each object (which can be edited using the Label view of the object in the workfile). Only one operator is available on these fields, the matches operator, which behaves exactly the same as the description field in the section on easy queries.

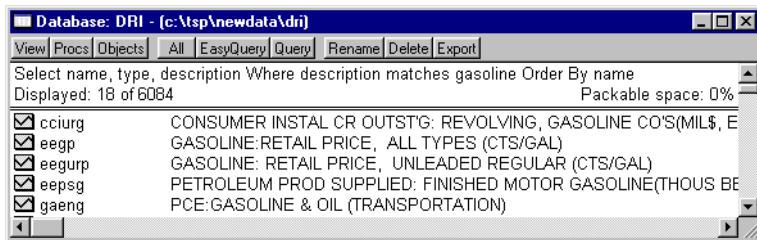
### Query Examples

Suppose you are looking for data related to gasoline consumption and gasoline prices in the database named DRIBASIC. First open the database: click **File/Open**, select **Files of type: Database .edb** and locate the database. From the database window, click **Query** and fill in the Advanced Query dialog as follows:

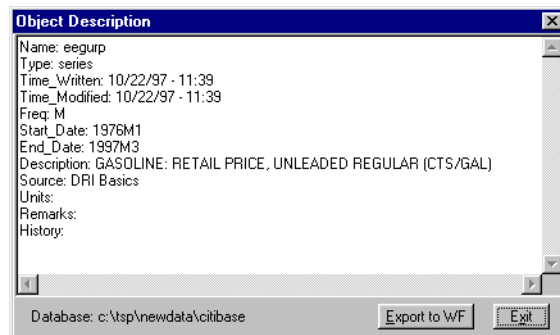
**Select:** name, type, description

**Where:** description matches gasoline

If there are any matches, the results are displayed in the database window like this:



To view the contents of all fields of an item, double click on its name. EViews will open an Object Description window that looks as follows:





To further restrict your search to series with at least quarterly frequency and to display the start and end dates of the results, click **Query** and again and modify the fields as follows:

**Select:** name, type, start, end, description

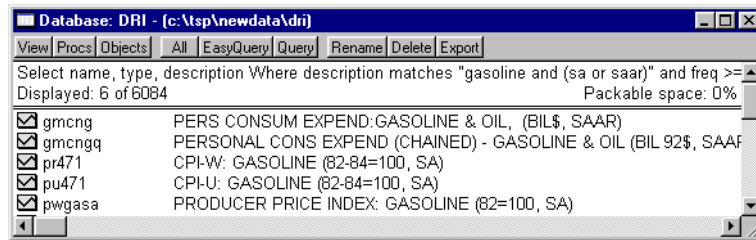
**Where:** description matches gasoline and freq>=q

If you are interested in seasonally adjusted series, which happen to contain sa or saar in their description in this database, further modify the fields to

**Select:** name, type, start, end, description

**Where:** description matches "gasoline and (sa or saar) " and freq>=q

The display of the query results now looks as follows:



## Maintaining the Database

In many cases a database should function adequately without any explicit maintenance. Where maintenance is necessary, EViews provides a number of procedures to help you perform common tasks.

### Database File Operations

Because databases are spread across multiple files, all of which have the same name but different extensions, simple file operations like copy, rename and delete require multiple actions if performed outside of EViews. The **Procs** button in the database window toolbar contains the procedures **Copy the database**, **Rename the database**, and **Delete the database** that carry out the chosen operation on all of the files that make up the database.

Note that file operations do not automatically update the database registry. If you delete or rename a database that is registered, you should either create a new database with the same name and location, or edit the registry.

### Packing the Database

If many objects are deleted from a database without new objects being inserted, a large amount of unused space will be left in the database. In addition, if objects are frequently overwritten in

the database, there will be a tendency for the database to grow gradually in size. The extent of growth will depend on circumstances, but a typical database is likely to stabilize at a size around 60% larger than what it would be if it were written in a single pass.

A database can be compacted down to its minimum size by using the pack procedure. Simply click on the button marked **Procs** in the toolbar at the top of the database window, then select the menu item **Pack the Database**. Depending on the size of the database and the speed of the computer which you are using, this operation may a significant amount of time.

You can get some idea of the amount of space that will be reclaimed during a pack by looking at the Packable Space percentage displayed in the top right corner of the database window. A figure of 30%, for example, indicates that roughly a third of the database file consists of unused space. A more precise figure can be obtained from the **Database Statistics** view of a database. The number following the label “unused space” gives the number of bytes of unused space contained in the main database file.

## Dealing with Errors

The best way to protect against damage to a database is to make regular backup copies of the database. This can be performed easily using the **Copy the Database** procedure documented above. EViews provides a number of other features to help you deal with damaged databases.

Damaged databases can be divided into two basic categories depending on how severely the database has been damaged. A database which can still be opened in a database window, but generates an error when performing some operations, may not be severely damaged, and may be able to be repaired. A database which can no longer be opened in a database window is severely damaged and will need to be rebuilt as a new database.

EViews has two procedures designed for working with databases which can be opened: **Test Database Integrity** and **Repair Database**. Both procedures are accessed by clicking on the button marked **Procs** in the database window toolbar, then selecting the appropriate menu item.

**Test Database Integrity** conducts a series of validity checks on the main database and index files. If an error is detected, a message box will be displayed providing some information as to the type of error found, and a suggestion as to how it might be dealt with. Because testing performs a large number of consistency checks on the database files, it may take considerable time to complete. You can monitor its progress by watching the messages displayed in the status line at the bottom of the EViews window. Testing a database does not modify the database in any way, and will never create additional damage to a database.

**Repair Database** will attempt to automatically detect and correct simple problems in the database. Although care has been taken to make this command as safe as possible, it will attempt to modify a damaged database, so it is probably best to make a back up copy of a damaged database before running this procedure.

If the database is badly corrupted, it is not possible for it to be repaired. In this case, EViews gives you the option of building a new database from the old one using the `dbrebuild` command. This operation can only be performed from the command line (since it may be impossible to open the database). The command is:

```
dbrebuild old_dbname new_dbname
```

The `dbrebuild` command does a low level scan through the main data file of the database `old_dbname` looking for any objects which can be recovered. Objects which it finds are copied into the new database `new_dbname`. This is a very time consuming process, but should be able to recover data from even heavily damaged files.

## Foreign Databases

### DRI Basic Economics Database

If you are a subscriber to the DRI Basic Economics database from Standard and Poor's / DRI, we have provided a way to quickly convert the entire database into the new EViews database format with only one command.

Simply type the following in the command window:

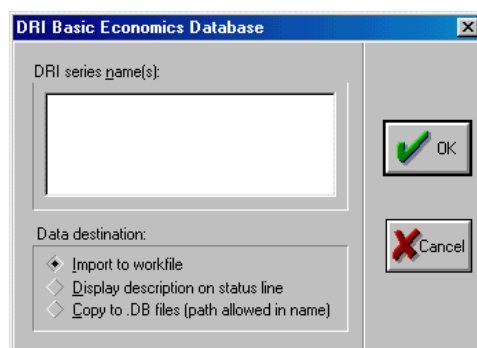
```
driconvert db_name
```

where `DB_NAME` is the name of a database which must already exist. This command will load every series in the DRI Basic Economics files and copy them into the database with name `DB_NAME`. Depending on the resources of your computer, the operation may take a significant amount of time. (On a 200MHz Pentium Pro, it takes around three minutes to construct the database of over 6000 series).

You may also access the DRI Basic Economics data directly without converting it to an EViews database. Click on **File/Import/DRI Basic Economics Database...** in the main menu.

Type in the names of the series as given in the DRI Basic Economics manual (you cannot query the unconverted DRI Basic Economics Database) in the blank area at the top of the dialog box. The buttons at the bottom let you choose one of three destinations for the data.

**Import to workfile** puts the series into the workfile, using the same names that the series have in the database. **Display description on status line** does not move the data, but just shows you the descriptive text from DRI at the bottom of your screen. **Copy to .DB files** puts each series



into a .DB file, using the DRI series name as the file name and .DB as the extension. With this choice, you may include a DOS path name to indicate which directory is the destination for the data.

Be sure to recopy TSPBSE.DIR into your EViews subdirectory every time you update the DRI Basic Economics database from the installation disks so that all of the updated series information may be found.

## Haver Analytics Databases

In EViews 3.1, similar conversion utility and direct access commands exist for Haver Analytics Databases.

The syntax of the Haver conversion command is

```
hconvert haver_db_path db_name
```

where HAVER\_DB\_PATH is the full path to the Haver database and DB\_NAME is the name of a database which must already exist. You can also fetch Haver series directly into EViews using the `hfetch` command.

## Connecting to Remote Databases

EViews has the ability to access databases on remote machines through an interface very similar to that provided for local EViews databases. In many cases, you can work with a remote database using exactly the same operations and commands you would use for a database on a local machine.

At the time of printing of this manual, remote access is only provided to databases hosted by Standard & Poor's/DRI (S&P/DRI). Subscribers to S&P/DRI data services can use these features to access data directly from within EViews. In the near future software will be available from QMS which will make it possible for anyone with an Internet or Intranet server to provide remote access to EViews format databases. Please contact our office for further information.

Although the interface to remote databases is very similar to that of local databases, there are some differences due to the nature of the connection. There are also some issues specifically related to accessing S&P/DRI data. The following sections document these differences.

### Enabling S&P / DRI Access

As of version 3.1, EViews is distributed with online access to S&P/DRI data services enabled. You need no longer contact QMS to activate the feature.

In order to access S&P/DRI data services, you will also need to have an active account with S&P/DRI. If you are not an existing S&P/DRI customer but may be interested in becoming one, you

should contact S&P/DRI for details (see the booklet included with EViews for contact information). S&P/DRI will often provide trial access to the service to interested customers.

Access to S&P/DRI data will not be possible unless you have already installed and configured the DRIPro server software. If you have not already done this, you should follow the instructions in the DRIPro booklet included with EViews. You should test that the DRIPro software is functioning correctly before attempting to use EViews to connect remotely to S&P/DRI. If you have difficulties with getting the S&P/DRI software to work, you should contact S&P/DRI directly for technical support.

## Creating a Database Link

A remote S&P/DRI database is represented in EViews by a *database link*. A database link resembles a local database, consisting of a set of files on disk, but instead of containing the data itself, a database link contains information as to how to access the remote data. A database link also contains a cache in which copies of recently retrieved objects are kept, which can substantially reduce the time taken to perform some database operations.

You can create a database link by following a similar procedure to that used to create a local database. Select **File/New/Database...** from the main menu, then, following the rules for local databases, enter the name you would like to give the new database link in the field **Database name or path:**. Then, click on the down arrow to the right of the field **Database type:** and select the option **Link to Remote S&P/DRI Database**. The dialog should change appearance so that a number of extra fields are displayed.

In the field **Connection name:** you should enter the name of the S&P/DRI connection you would like to use, as it appears in the **Connection Settings** box in the DRIPro configuration program. If you have only configured a single connection, and have not modified the connection name, the connection name will be DEFAULT, and this will be filled in automatically by EViews if you leave the field blank.

In the field **S&P/DRI Databank:** you should input the full name of the S&P/DRI bank to which you would like to connect, not including any leading @ sign. For example, to connect to the DRI Basic Economics Foreign Exchange database, you should enter “basic/fx”. (See the separate DRIPro booklet for a full list of DRI/S&P banks and their names). Each EViews database link can be associated with only one S&P/DRI databank, although you can create as many database links as you require.

The screenshot shows the 'Eviews Database' dialog box. It has a title bar with the text 'Eviews Database' and a close button. The dialog contains the following elements:

- Database name or path:** A text input field.
- Database type:** A dropdown menu currently showing 'Link to Remote DRI Database'. Below it is a checkbox labeled 'Add to Database Registry'.
- DRI link characteristics:** A section containing three input fields:
  - Connection name:**
  - DRI Databank:**
  - Local Password:**
- Buttons:** At the bottom, there are two buttons: 'OK' with a green checkmark icon and 'Cancel' with a red X icon.

The field **Local Password:** can be used to set a password that must be entered whenever you wish to use the database link. This should not be confused with your S&P/DRI username and password, which you must already have provided in the S&P/DRI configuration program. Accessing a database link which contains a local password will cause a dialog to appear which prompts the user to input the password. Access to the remote database is only provided if the remote password is valid. Leave this field blank if you do not want a password to be attached to the database link.

When you have finished filling in the dialog fields, click on the OK button. A new database will be created and a database window should appear on the screen.

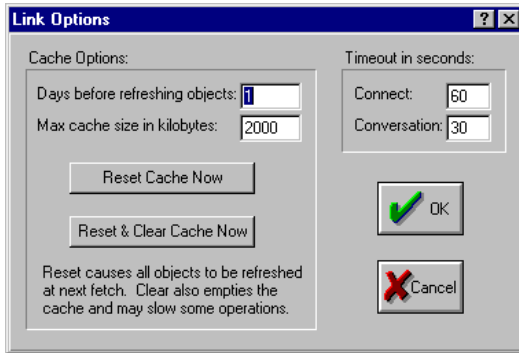
The S&P/DRI database link window is very similar to a normal EViews database window. You should be able to perform basic query operations and simple fetching of series without any special instructions. Note, however, that it is not possible to modify a remote S&P/DRI database from within EViews, so operations which involve writing to the database have been removed. There are a number of other complications related to dealing with S&P/DRI databases, and we discuss these in detail in the sections below.

## Understanding the Cache

A database link includes a cache of recently fetched objects which is used to speed up certain operations on the database. In some circumstances, fetching an object from the database will simply retrieve a copy from the local cache, rather than fetching a fresh copy of the data from the remote site. Even if a fresh copy is retrieved, having a previous copy of the series in the cache can substantially speed up retrieval.

You can regulate the caching behavior of the database link in a number of different ways. The basic option which determines under what circumstances a new copy of the data should be fetched is the *days before refresh*. If you attempt to fetch an object from the database link, and the copy of the object currently in the cache was fetched more recently than the days before refresh value, then the object currently in the cache will be returned instead of a fresh copy being fetched. For example, if days before refresh is set to one, any object which has already been fetched today will be retrieved from the cache, while any object which has not yet been fetched today will be retrieved from the remote site. Similarly, if days before refresh is set to seven, then an object in the cache must be more than a week old before a new copy of the object will be fetched. If days before refresh is set to zero, then a new copy of the data is fetched every time it is used.

You can change the days before refresh setting by clicking on the **Procs** button at the top of the database link window, then choosing **Link Options...** from the pop-up menu. A dialog will appear:



The dialog contains a number of fields, one of which is labeled **Days before refreshing objects**. Type a new number in the field to change the value.

The same dialog also contains a button marked **Reset cache now**. This button can be used to modify the behavior documented above. Clicking on the button causes the cache to mark all objects in the cache as out of date, so that the next time each object is fetched, it is guaranteed that a fresh copy will be retrieved. This provides a simple way for

you to be certain that the database link will not return any data fetched before a particular time.

The dialog also contains some options for managing the size of the cache. The field marked **Maximum cache size in kilobytes** can be used to set the maximum size that the cache will be allowed to grow to on disk. If the cache grows above this size, a prompt will appear warning you that the cache has exceeded the limit and asking if you would like to compact the cache. Compacting is performed by deleting objects from oldest to newest until the cache size is reduced to less than three quarters of its maximum size. The cache is then packed to reclaim the empty space.

You can also completely clear the contents of the cache at any time by clicking on the button marked **Reset & Clear Cache Now**.

You can always examine the current contents of the database cache by clicking on the **Cache** button at the top of the database link window. This will display the names of all objects currently in the cache.

## Configuring Link Options

The link options dialog also allows you to specify a number of timeout values. In most cases the default values will behave acceptably. If you believe you are having problems with EViews aborting the connection too early, or you would like to shorten the times so as to receive a timeout message sooner, then enter in new values in the appropriate fields.

**Connection timeout**—is the length of time, in seconds, that EViews will wait for a response when first connecting to S&P/DRI. Depending on the type of connection you are making to S&P/DRI, this can take a significant amount of time.

**Conversation timeout**—is the length of time, in seconds, that EViews will wait for a response from S&P/DRI when carrying out a transaction after a connection has already been made.

The values are attached to a particular database link, and can be reset at any time.

## Dealing with Illegal Names

S&P/DRI databanks contain a number of series with names which are not legal names for EViews objects. In particular, S&P/DRI names frequently contain the symbols @, & and %, none of which are legal characters in EViews object names. We have provided a number of features to allow you to work with these series within EViews.

Because the @ symbol is so common in S&P/DRI names, while the underline symbol (which is a legal character in EViews) is unused, we have hard-coded the rule that all underlines in EViews are mapped into @ symbols in S&P/DRI names when performing operations on an S&P/DRI database link. For example, if there is a series which S&P/DRI documentation suggests has the name JQIMET@UK, you should refer to this series inside EViews as JQIMET\_UK. Note that when performing queries, EViews will automatically replace the @ symbol by an underline in the object name before displaying the query results on the screen. Consequently, if you are fetching data by copying and pasting objects from a query window, you do not need to be aware of this translation.

For other names which are illegal in EViews (either because they contain illegal characters, or because they conflict with one of EViews reserved words), we have provided a very general facility which allows you to create a permanent mapping from the illegal names in a particular database into legal EViews names of your own choice. The basic idea is as follows: each database can have one or more *object aliases* associated with it where each alias entry consists of the name of the object in the database and the name by which you would like it to be known in EViews.

The easiest way to create an object alias is simply to try and fetch the object with the illegal name into EViews. If you are working with query results, you can tell which object names are illegal because they will be displayed in the database window in red. When you try to fetch an object with an illegal name, a dialog will appear.



The field labeled **EViews Name:** initially contains the illegal name of the database object. You should edit this to form a legal EViews object name. In this example, we could change the name GNP%N\_BR to GNP\_N\_BR. The checkbox labeled **Add this name to the database alias list** (which is checked by default), determines whether you want to create a permanent association between the name you have just typed and the illegal name. If you leave the box checked, then whenever you use the edited object name in the future, EViews will take it to refer to the underlying illegal name. The edited name acts as an *alias* for the underlying name. It is as though you had renamed the object in the database to the new legal name, except that you have not actually modified the database itself, and your changes will not affect other users of the database. You can uncheck the box if you wish to fetch

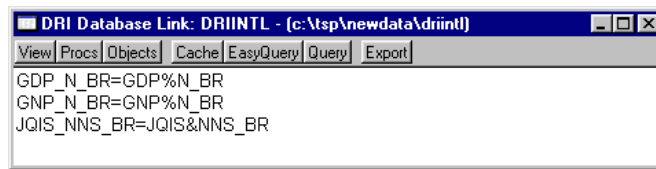


the object under a different name, but do not want to create any permanent association between the two names.

When EVIEWS displays an object in the database window for which an alias has been set, EVIEWS will show the alias, rather than the underlying name of the object. In order to indicate that this substitution has been done, EVIEWS displays the alias in blue.

Creating an alias can cause *shadowing* of object names. Shadowing occurs when you create an alias for an object in the database, but the name you use as an alias is the name of another object in the database. Because the existence of the alias will stop you from accessing the other object, that object is said to be shadowed. In order to show you that an object name being display has been shadowed, EVIEWS displays the name of shadowed objects in green. You will not be able to fetch an object which has been shadowed without modifying either its name or the alias which is causing it to be shadowed. Even if the shadowed series is explicitly selected with the mouse, operations performed on the series will use the series with the conflicting alias, not the shadowed series.

You can view a list of the aliases currently defined for any database by clicking on the **View** button at the top of the database window, then selecting **Object Aliases**. A list of all the aliases will be displayed in the window.



Each line represents one alias attached to the database and follows the format:

```
alias=database_object_name
```

You can edit the list of aliases to delete unwanted entries, or you can type in, or cut and paste, new entries into the file. You must follow the rule that both the set of aliases and the set of database names do not contain any repeated entries. (If you do not follow this rule, EVIEWS will refuse to save your changes). To save any modifications you have made, simply switch back to the **Object Display** view of the database. EVIEWS will prompt you for whether you want to save or discard your edits.

The list of currently defined database aliases for all databases is kept in the file OBALIAS.INI in the EVIEWS installation directory. If you would like to replicate a particular set of aliases onto a different machine, you should copy this file to the other machine, or use a text editor to combine a portion of this file with the file already in use on the other machine. You must exit and restart EVIEWS to be sure that EVIEWS will reread the aliases from the file.

## Issues with S&P / DRI Frequencies

S&P/DRI databases have a different structure from EViews databases. An EViews database can contain series with mixed frequencies. A S&P/DRI database can contain data of only a single frequency. In order that similar data may be grouped together, each S&P/DRI databank is actually composed of a series of separate databases, one for each frequency. When working with S&P/DRI data from within S&P/DRI software, you will often have to specify at exactly which frequency a particular series can be found. In some cases, an S&P/DRI databank may contain a series with the same name stored at several different frequencies.

Because this approach is inconsistent with the way that EViews own databases work, we have tried to create a simpler interface to S&P/DRI data where you do not need to keep track of the frequency of each series that you would like to fetch. Instead, you can simply fetch a series by name or by selecting it from the query window, and EViews will do whatever is necessary to find out the frequency for you.

An ambiguity can arise in doing this, where a series with the same name appears at a variety of different frequencies in the S&P/DRI databank. By default, EViews resolves this ambiguity by always fetching the *highest* frequency data available. EViews will then perform necessary frequency conversions using the standard rules for frequency conversion in EViews (see “Frequency Conversion” above).

In many cases, this procedure will exactly replicate the results that would be obtained if the lower frequency data was fetched directly from S&P/DRI. In some cases (typically when the series in question is some sort of ratio or other expression of one or more series) the figures may not match up exactly. In this case, if you know that the S&P/DRI data exists at multiple frequencies and you are familiar with S&P/DRI frequency naming conventions, you can explicitly fetch a series from S&P/DRI at a particular frequency by using a modified form of the command line form of fetch. Simply add the S&P/DRI frequency in parentheses after the name of the series. For example, the command:

```
fetch x(Q) y(A)
```

will fetch the series *x* and *y* from the current default database, reading the quarterly frequency copy of *x* and the annual frequency copy of *y*. If you request a frequency at which the data is not available, you will receive an error message. You should consult S&P/DRI documentation for details on S&P/DRI frequencies.

## Limitations of S&P / DRI Queries

Queries to S&P/DRI database links are more limited than those available for EViews databases. The following section documents the restrictions.

First, queries on S&P/DRI databases allow only a subset of the fields available in EViews databases to be selected. The fields supported are: name, type, freq, start, end, last\_update and description.

Second, the only fields which can be used in ‘where’ conditions in a query on a S&P/DRI database link are name and description. (EViews does not support queries by frequency because of the ambiguities arising from S&P/DRI frequencies noted above).

Each of these fields has only one operator, the ‘matches’ operator, and operations on the two fields can only be joined together using the ‘and’ operator.

The ‘matches’ operator is also limited for queries on S&P/DRI databases, matching only a subset of the expressions available for EViews databases. In particular, the pattern expression in a query on an S&P/DRI database must either have the form

a or b or ... c

or the form

a and b and ... c

Mixing of ‘and’ and ‘or’ is not allowed, and the ‘not’ operator is not supported.

Patterns, however, are allowed and follow the normal EViews rules where ? denotes any single character and \* denotes zero or more characters.

Sorting of results by field is not supported.

## Dealing with Common Problems

As stated in the introduction, you must install and configure the S&P/DRI software before EViews will be able to connect to S&P/DRI. If you cannot connect to S&P/DRI using the DRIPRO software, you should contact S&P/DRI directly for assistance.

Assuming that you have correctly configured your S&P/DRI connection, in most cases EViews will be able to recover adequately from unexpected problems which arise during a S&P/DRI session without user intervention. Sometimes this will require EViews to automatically disconnect then reconnect to S&P/DRI.

There are some circumstances in which EViews may have problems making a connection. In order to connect to S&P/DRI, EViews uses a program written by S&P/DRI called DRIPROSV. You can tell when this program is running by looking for the icon labeled “DRIPRO server” in the Windows taskbar. Because of problems that can arise with multiple connections, EViews will not attempt to use the program if it is already running. Instead, EViews will report an error message “DRI server software already running”. If there is another application which is using the connection to S&P/DRI, you can simply close down that program and the S&P/DRI server software

should shut down automatically. If this is not the case, you may have to close down the S&P/DRI server software manually. Simply click on the icon in the Windows taskbar with the right mouse button, then select **Close** from the pop-up menu.

You can also use this as a procedure for forcing the S&P/DRI connection to terminate. (For example to hang up the phone if you are using a direct modem rather than an Internet connection). Closing down the server software may cause EViews to report an error if it is currently carrying out a database transaction, but should otherwise be safe. EViews will restart the server software whenever it is needed.

Note that running other DRIPro software while EViews is using the S&P/DRI server software may cause EViews to behave unreliably.

## Commands

To create a new database named USDB, type

```
dbcreate c:\evdata\usdb
```

To open the database DRIBASIC and make it the default database, type

```
dbopen dribasic
```

To open, if not already open, or to create a new database if one does not already exist, a database named US1, type

```
db us1
```

Note that this command also makes US1 the default database.

To make a backup copy US1\_BAK of the entire database US1, type

```
dbcopy us1 us1_bak
```

To delete the entire US1 database, type

```
dbdelete us1
```

To store objects (named GDP and M1) in the default database, type

```
store gdp m1
```

To store the object named GDP into two different databases (named US1 and FINANCE1), type

```
store us1::gdp finance1::gdp
```

To store GDP into the default database and all objects whose name begins with the letter X into the database named US1, type

```
store(1) gdp us1::x*
```

The option 1 in parentheses instructs EViews to save space by storing all series in single rather than double precision.

To store two objects (GDP and M1) into individual .DB files, type

```
store(i) gdp c:\evdata\m1
```

To fetch two objects GDP and M1, searching all databases in the search order, type

```
fetch(d) gdp m1
```

To fetch two objects from individual .DB files, with frequency conversion by summation if necessary, type

```
fetch(i,c=s) gdp m1
```

Fetch all objects that have GDP in their names from two databases, with frequency conversion by averaging if necessary, type

```
fetch(c=a) us1::*gdp* dri::*gdp*
```

To make a backup copy of GDP within the US1 database, type

```
copy us1::gdp us1::gdp_bak
```

To copy all objects in the US1 database to another existing database WORLD1, type

```
copy us1::* world1::
```



## Part II. Basic Data Analysis

---

Chapters 7–10 describe the EViews objects that you will use to perform basic data analysis.

- Chapter 7 describes the series object. Series are the basic unit of data in EViews and are the basis for all univariate analysis. This chapter documents the basic graphing and data analysis features associated with series.
- Chapter 8 documents the group object. Groups are collections of series which form the basis for a variety of multivariate graphing and data analyses.
- Chapter 9 provides detailed documentation for exploratory data analysis using distribution graphs, kernel density, and scatterplot fit graphs.
- Chapter 10 describes the creation and customization of tables and graphs.





## Chapter 7. Series

---

EViews provides various statistical graphs, descriptive statistics, and procedures as *views* and *procedures* of a series. Once you have read or generated data into series objects using any of the methods described in Chapters 4–6, you are ready to perform statistical and graphical analysis using the data contained in the series.

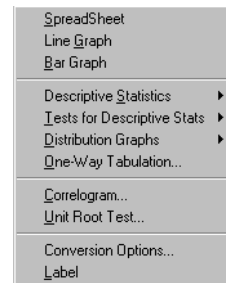
Series views compute various statistics for a single series and display these statistics in various forms such as spreadsheets, tables, and graphs. The views range from a simple line graph, to kernel density estimators. Series procedures create new series from the data in existing series. These procedures include various seasonal adjustment methods, exponential smoothing methods, and the Hodrick-Prescott filter.

Methods which involve more than one series are described in Chapter 8, “Groups”, which outlines views and procedures for group objects.

To access the views and procedures for series, open the series window by double clicking on the series name in the workfile, or by typing `show` followed by the name of the series in the command window.

### Series Views

The series view drop-down menu is divided into four blocks. The first block lists views of the underlying data in the series. The second and third blocks provide access to general statistics; the views in the third block are mainly for time series. The fourth block contains the label view.



#### Spreadsheet, Line Graph, and Bar Graph Views

- **Spreadsheet** displays the raw data in the series in spreadsheet format.
- **Line Graph** plots the series against the date/observation number. This view is most useful for time series. See Chapter 10 for a discussion of techniques for modifying and customizing the graphical display.
- **Bar Graph** plots the bar graph of the series. This view is useful for plotting series from a small data set that takes only a few distinct values.

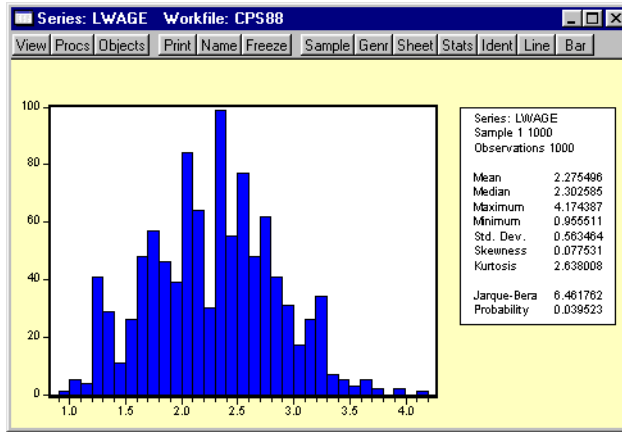
## Descriptive Statistics

This view displays various summary statistics for the series.

### Histogram and Stats

This view displays the frequency distribution of your series in a histogram. The histogram divides the series range (the distance between the maximum and minimum values) into a number of equal length intervals or bins and displays a count of the number of observations that fall into each bin.

Histogram and Stats  
Stats by Classification...



A complement of standard descriptive statistics are displayed along with the histogram. All of the statistics are calculated using observations in the current sample.

**Mean** is the average value of the series, obtained by adding up the series and dividing by the number of observations.

**Median** is the middle value (or average of the two middle values) of the series when the values are ordered from the smallest to the largest. The median is a robust measure of the center of the distribution that is less sensitive to outliers than the mean.

**Max** and **Min** are the maximum and minimum values of the series in the current sample.

**Std. Dev.** (standard deviation) is a measure of dispersion or spread in the series. The standard deviation is given by:

$$\hat{s} = \sqrt{\left( \sum_{i=1}^N (y_i - \bar{y})^2 \right) / (N-1)} \quad (7.1)$$

where  $N$  is the number of observations in the current sample and  $\bar{y}$  is the mean of the series.

**Skewness** is a measure of asymmetry of the distribution of the series around its mean. Skewness is computed as:

$$S = \frac{1}{N} \sum_{i=1}^N \left( \frac{y_i - \bar{y}}{\hat{\sigma}} \right)^3 \quad (7.2)$$

where  $\hat{\sigma}$  is an estimator for the standard deviation that is based on the biased estimator for the variance ( $\hat{\sigma} = \hat{s}\sqrt{(N-1)/N}$ ). The skewness of a symmetric distribution, such as the normal distribution, is zero. Positive skewness means that the distribution has a long right tail and negative skewness implies that the distribution has a long left tail.

**Kurtosis** measures the peakedness or flatness of the distribution of the series. Kurtosis is computed as

$$K = \frac{1}{N} \sum_{i=1}^N \left( \frac{y_i - \bar{y}}{\hat{\sigma}} \right)^4 \quad (7.3)$$

where  $\hat{\sigma}$  is again based on the biased estimator for the variance. The kurtosis of the normal distribution is 3. If the kurtosis exceeds 3, the distribution is peaked (leptokurtic) relative to the normal; if the kurtosis is less than 3, the distribution is flat (platykurtic) relative to the normal.

**Jarque-Bera** is a test statistic for testing whether the series is normally distributed. The test statistic measures the difference of the skewness and kurtosis of the series with those from the normal distribution. The statistic is computed as:

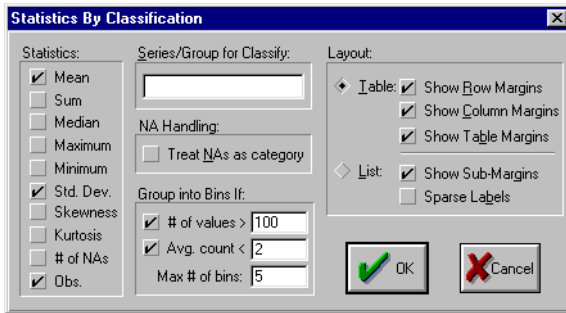
$$\text{Jarque-Bera} = \frac{N-k}{6} \left( S^2 + \frac{(K-3)^2}{4} \right) \quad (7.4)$$

where  $S$  is the skewness,  $K$  is the kurtosis, and  $k$  represents the number of estimated coefficients used to create the series.

Under the null hypothesis of a normal distribution, the Jarque-Bera statistic is distributed as  $\chi^2$  with 2 degrees of freedom. The reported Probability is the probability that a Jarque-Bera statistic exceeds (in absolute value) the observed value under the null—a small probability value leads to the rejection of the null hypothesis of a normal distribution. For the LWAGE series displayed above, we reject the hypothesis of normal distribution at the 5% level but not at the 1% significance level.

## Stats by Classification

This view allows you to compute the descriptive statistics of a series for various subgroups of your sample. If you select **View/Descriptive Statistics/Stats by Classification...**, a Statistics by Classification dialog box appears:



The **Statistics** option at the left allows you to choose the statistics you wish to compute.

In the **Series/Group for Classify** field enter series or group names that define your subgroups. You must type at least one name. Descriptive statistics will be calculated for each unique value of the classification series unless binning is selected. You may type more than one series or group name; separate each name by a space.

By default, EViews excludes observations which have missing values for any of the classification series. To treat NA values as a valid subgroup, select the **NA handling** option.

The **Layout** option allows you to control the display of the statistics. Table layout arrays the statistics in cells of two-way tables. The list form displays the statistics in a single line for each classification group. We provide examples of both formats below.

The Table and List options are only relevant if you use more than one series as a classifier. The **Sparse Labels** option suppresses repeating labels in list mode to make the display less cluttered.

The **Row Margins**, **Column Margins**, and **Table Margins** instruct EViews to compute statistics for aggregates of your subgroups. For example, if you classify your sample on the basis of gender and age, EViews will compute the statistics for each gender/age combination. If you elect to compute the marginal statistics, EViews will also compute statistics corresponding to each gender, and each age subgroup.

A classification may result in a large number of distinct values with very small cell sizes. By default, EViews automatically groups observations to maintain moderate cell sizes and numbers of categories. **Group into Bins** provides you with control over this process:

Setting the **# of values** option bins tell EViews to group data if the classifier series takes more than the specified number of distinct values.

The **Avg. count** option bins the series if the average count for each distinct value of the classifier series is less than the specified number.

The **Max # of bins** specifies the maximum number of subgroups to bin the series. Note that this number only provides you with approximate control over the number of bins.

The default setting is to bin the series into 5 subgroups if either the series takes more than 100 distinct values or if the average count is less than 2. If you do not want to bin the series, unmark both options.

For example, consider the following stats by classification view in table form:

Descriptive Statistics for LWAGE  
 Categorized by values of MARRIED and UNION  
 Date: 10/15/97 Time: 01:11  
 Sample: 1 1000  
 Included observations: 1000

Mean Median Std. Dev. Obs.		UNION		
		0	1	All
0		1.993829	2.387019	2.052972
		1.906575	2.409131	2.014903
		0.574636	0.395838	0.568689
		305	54	359
MARRIED 1		2.368924	2.492371	2.400123
		2.327278	2.525729	2.397895
		0.557405	0.380441	0.520910
		479	162	641
All		2.223001	2.466033	2.275496
		2.197225	2.500525	2.302585
		0.592757	0.386134	0.563464
		784	216	1000

The header indicates that the table cells are categorized by two series MARRIED and UNION. These two series are dummy variables that take only two values and no binning was made. If the series were binned, intervals rather than a number will be displayed in the margins.

The upper left cell of the table indicates the reported statistics in each cell; in this case the median and the number of observations are reported in each cell. The row and column labeled **All** correspond to the **Row Margin** and **Column Margin** options described above.

The same view in list form with sparse labels is displayed below:

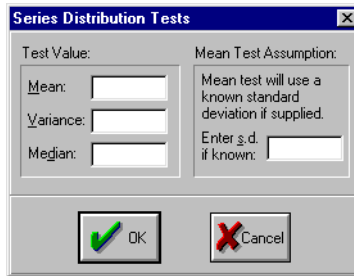
Descriptive Statistics for LWAGE  
 Categorized by values of MARRIED and UNION  
 Date: 10/15/97 Time: 01:08  
 Sample: 1 1000  
 Included observations: 1000

UNION	MARRIED	Mean	Median	Std. Dev.	Obs.
0	0	1.993829	1.906575	0.574636	305
	1	2.368924	2.327278	0.557405	479
	All	2.223001	2.197225	0.592757	784
1	0	2.387019	2.409131	0.395838	54
	1	2.492371	2.525729	0.380441	162
	All	2.466033	2.500525	0.386134	216
All	0	2.052972	2.014903	0.568689	359
	1	2.400123	2.397895	0.520910	641
	All	2.275496	2.302585	0.563464	1000

## Tests for Descriptive Stats

### Simple Hypothesis Tests

This view carries out simple hypothesis tests regarding the mean, median, and the variance of the series. These are all single sample tests; see Equality Tests by Classification below for two sample tests. If you select **View/Tests for Descriptive Stats/Simple Hypothesis Tests**, the Series Distribution Tests dialog box appears



### Mean Test

Carries out the test of the null hypothesis that the mean  $\mu$  of the series  $x$  is equal to a specified value  $m$  against the two-sided alternative that it is not equal to  $m$ :

$$\begin{aligned} H_0: \mu &= m \\ H_1: \mu &\neq m \end{aligned} \quad (7.5)$$

If you do not specify the standard deviation of  $x$ , EViews reports a  $t$ -statistic based on an estimated standard deviation of  $x$  computed as:

$$t = \frac{\bar{x} - m}{s/\sqrt{N}} \quad (7.6)$$

where  $\bar{x}$  is the sample mean of  $x$  and  $N$  is the number of observations of  $x$ . Under the null hypothesis and if  $x$  is normally distributed, the  $t$ -statistic has a  $t$ -distribution with  $N - 1$  degrees of freedom.

If you specify a value for the standard deviation of  $x$ , EViews also reports a  $z$ -statistic

$$z = \frac{\bar{x} - m}{\sigma/\sqrt{N}} \quad (7.7)$$

where  $\sigma$  is the specified standard deviation of  $x$ . Under the null hypothesis and if  $x$  is normally distributed with standard deviation  $\sigma$ , the  $z$ -statistic has a standard normal distribution.

To carry out the mean test, type in the value of the mean under the null hypothesis in the edit field next to **Mean**. If you want to compute the  $z$ -statistic conditional on a known standard devi-

ation, also type in a value for the standard deviation in the right edit field. You can type in any number or standard EViews expression in the edit fields. Below, we present the output of this test:

Hypothesis Testing for LWAGE		
Date: 10/15/97 Time: 01:14		
Sample: 1 1000		
Included observations: 1000		
<hr/>		
Test of Hypothesis: Mean = 2		
<hr/>		
Sample Mean = 2.275496		
Sample Std. Dev. = 0.563464		
<hr/>		
<u>Method</u>	<u>Value</u>	<u>Probability</u>
t-statistic	15.46139	0.00000
<hr/>		

The reported probability value is the  $p$ -value, or marginal significance level, against a two-sided alternative. If this probability value is less than the size of the test, say 0.05, we reject the null hypothesis. Here, we strongly reject the null hypothesis for the two-sided test of equality. The probability value for a one-sided alternative is one half the  $p$ -value of the two-sided test.

### Variance Test

Carries out the test of the null hypothesis that the variance of a series  $X$  is equal to a specified value  $\sigma^2$  against the two-sided alternative that it is not equal to  $\sigma^2$ :

$$\begin{aligned} H_0: \text{var}(x) &= \sigma^2 \\ H_1: \text{var}(x) &\neq \sigma^2 \end{aligned} \quad (7.8)$$

EViews reports a “ $\chi^2$ -statistic” computed as:

$$\chi^2 = \frac{(N-1)s^2}{\sigma^2} \quad (7.9)$$

where  $N$  is the number of observations and  $\bar{x}$  is the sample mean of  $x$ . Under the null hypothesis and the assumption that  $X$  is normally distributed, the  $\chi^2$ -statistic has a  $\chi^2$ -distribution with  $N-1$  degrees of freedom. The probability value is computed as  $\min(p, 1-p)$ , where  $p$  is the probability of observing a  $\chi^2$ -statistic as large as the one actually observed under the null hypothesis.

To carry out the variance test, type in the value of the variance under the null hypothesis in the field box next to Variance. You can type in any *positive* number or expression in the field.

### Median Test

Carries out the test of the null hypothesis that the median of a series  $x$  is equal to a specified value  $m$  against the two-sided alternative that it is not equal to  $m$ :

$$\begin{aligned}H_0: \text{med}(x) &= m \\ H_1: \text{med}(x) &\neq m\end{aligned}\tag{7.10}$$

EViews reports three rank-based, nonparametric test statistics. The principal references for this material are Conover (1980) and Sheskin (1997).

**Binomial sign test.** This test is based on the idea that if the sample is drawn randomly from a binomial distribution, the sample proportion above and below the true median should be one-half. Note that EViews reports two-sided  $p$ -values for both the sign test and the large sample normal approximation (with continuity correction).

**Wilcoxon signed-ranks test.** Suppose that we compute the absolute value of the difference between each observation and the mean, and then rank these observations from high to low. The Wilcoxon test is based on the idea that the sum of the ranks for the samples above and below the median should be similar. EViews reports a  $p$ -value for the normal approximation to the Wilcoxon  $T$ -statistic (correcting for both continuity and ties).

**Van der Waerden (normal scores) test.** This test is based on the same general idea as the Wilcoxon test, but is based on smoothed ranks. The signed ranks are smoothed by converting them to quantiles of the normal distribution (normal scores). EViews reports the two-sided  $p$ -value for the asymptotic normal test described by Conover (1980).

To carry out the median test, type in the value of the median under the null hypothesis in the edit box next to **Median**. You can type any numeric expression in the edit field.



## Hypothesis Testing for LWAGE

Date: 10/14/97 Time: 23:23

Sample: 1 1000

Included observations: 1000

Test of Hypothesis: Median = 2.25

Sample Median = 2.302585

Method	Value	Probability
Sign (exact binomial)	532	0.046291
Sign (normal approximation)	1.992235	0.046345
Wilcoxon signed rank	1.134568	0.256556
van der Waerden (normal scores)	1.345613	0.178427

## Median Test Summary

Category	Count	Mean Rank
Obs > 2.25	532	489.877820
Obs < 2.25	468	512.574786
Obs = 2.25	0	
Total	1000	

## Equality Tests by Classification

This view allows you to test equality of the means, medians, and variances across subsamples (or subgroups) of a single series. For example, you can test whether mean income is the same for males and females, or whether the variance of education is related to race. The tests assume that the subsamples are independent.

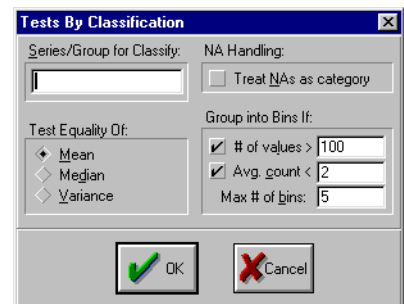
For single sample tests, see the discussion of “Simple Hypothesis Tests” on page 148, above. For tests of equality across different series, see “Equality Tests by Classification” on page 151.

Select **View/Tests for Descriptive Stats/Equality Tests by Classification....** The Tests by Classification dialog box will appear.

First, select whether you wish to test the mean, the median or the variance. Specify the subgroups, the NA handling, and the grouping options as described above in the discussion for “Statistics by Classification” beginning on page 146.

### Mean Equality Test

This test is based on a single-factor, between-subjects, analysis of variance (ANOVA). The basic idea is that if the subgroups have the same mean, then the variability between the sample means (between group) should be the same as the variability within any subgroup (within group).



Denote the  $i$ -th observation in group  $g$  as  $x_{g,i}$ , where  $i = 1, \dots, n_g$  for groups  $g = 1, \dots, G$ . The between and within sums of squares are defined as

$$SS_B = \sum_{g=1}^G n_g (\bar{x}_g - \bar{x})^2 \quad (7.11)$$

$$SS_W = \sum_{g=1}^G \sum_{i=1}^{n_g} (x_{ig} - \bar{x}_g)^2 \quad (7.12)$$

where  $\bar{x}_g$  is the sample mean within group  $g$  and  $\bar{x}$  is the overall sample mean. The  $F$ -statistic for the equality of means is computed as

$$F = \frac{SS_B/(G-1)}{SS_W/(N-G)} \quad (7.13)$$

where  $N$  is the total number of observations. The  $F$ -statistic has an  $F$ -distribution with  $G-1$  numerator degrees of freedom and  $N-G$  denominator degrees of freedom under the null hypothesis of independent and identical normal distribution, with equal means and variances in each subgroup.

For tests with only two subgroups ( $G = 2$ ), EViews also reports the  $t$ -statistic, which is simply the square root of the  $F$ -statistic with one numerator degree of freedom:

Test for Equality of Means of LWAGE  
 Categorized by values of MARRIED and UNION  
 Date: 10/14/97 Time: 12:28  
 Sample: 1 1000  
 Included observations: 1000

Method	df	Value	Probability
Anova F-statistic	(3, 996)	43.40185	0.0000

#### Analysis of Variance

Source of Variation	df	Sum of Sq.	Mean Sq.
Between	3	36.66990	12.22330
Within	996	280.5043	0.281631
Total	999	317.1742	0.317492

#### Category Statistics

UNION	MARRIED	Count	Mean	Std. Dev.	Std. Err. of Mean
0	0	305	1.993829	0.574636	0.032904
0	1	479	2.368924	0.557405	0.025468
1	0	54	2.387019	0.395838	0.053867
1	1	162	2.492371	0.380441	0.029890
All		1000	2.275496	0.563464	0.017818

The analysis of variance table shows the decomposition of the total sum of squares into the between and within sum of squares, where

$$\text{Mean Sq.} = \text{Sum of Sq.} / \text{df}$$

The  $F$ -statistic is the ratio

$$F = \text{Between Mean Sq.} / \text{Within Mean Sq.}$$

### Median Equality Tests

EViews computes various rank-based nonparametric tests of the hypothesis that the  $G$  subgroups have the same median, against the alternative that at least one subgroup has a different median.

**Mann-Whitney  $U$ -test.** This test statistic is only reported when there are two subgroups ( $G = 2$ ). The test is similar in spirit to the Wilcoxon test outlined above on page Wilcoxon signed-ranks test. Suppose that we compute the absolute value of the difference between each observation and the mean, and then rank these observations from high to low. The Wilcoxon test is based on the idea that the sum of the ranks for the samples above and

below the median should be similar. EViews reports a  $p$ -value for the normal approximation to the Wilcoxon T -statistic (correcting for both continuity and ties).. The idea is to rank the series from smallest value (rank 1) to largest, and to compare the sum of the ranks from subgroup 1 to the sum of the ranks from subgroup 2. If the groups have the same median, the values should be similar. EViews reports the normal approximation to the  $U$ -statistic (with continuity and tie correction) and the  $p$ -values for a two-sided test. For details, see Sheskin (1997).

**Kruskal-Wallis one-way ANOVA by ranks.** This is a generalization of the Mann-Whitney test to more than two subgroups. The test is based on a one-way analysis of variance using only ranks of the data. EViews reports the  $\chi^2$  chi-square approximation to the Kruskal-Wallis test statistic (with tie correction). Under the null hypothesis, this statistic is approximately distributed as a  $\chi^2$  with  $G - 1$  degrees of freedom (see Sheskin, 1997).

**van der Waerden (normal scores) test.** This test is analogous to the Kruskal-Wallis test, except what we smooth the ranks by converting them into normal quantiles (Conover, 1980). EViews reports a statistic which is approximately distributed as a  $\chi^2$  with  $G - 1$  degrees of freedom under the null hypothesis.

**Chi-square test for the median.** This is a rank-based ANOVA test based on the comparison of the number of observations above and below the overall median in each subgroup. This test is also known as the median test (Conover, 1980)).

Under the null hypothesis, the median chi-square statistic is asymptotically distributed as a  $\chi^2$  with  $G - 1$  degrees of freedom. EViews also reports Yates' continuity corrected statistic. You should note that the use of this correction is controversial (Sheskin, 1997, p. 218).

In addition to the test statistics and  $p$ -values, EViews reports values for the components of the test statistics for each subgroup of the sample. For example, the column labeled **Mean Score** contains the mean values of the van der Waerden scores (the smoothed ranks) for each subgroup.

### Variance Equality Tests

Tests the null hypothesis that the variances in all  $G$  subgroups are equal against the alternative that at least one subgroup has a different variance. See Conover, *et al.* (1981) for a general discussion of variance testing.

**$F$ -test.** This test statistic is reported only for tests with two subgroups ( $G = 2$ ). Compute the variance for each subgroup and denote the subgroup with the larger variance as  $L$  and the subgroup with the smaller variance as  $S$ . Then the  $F$ -statistic is given by

$$F = s_L^2 / s_S^2 \quad (7.14)$$

where  $s_g^2$  is the variance in subgroup  $g$ . This  $F$ -statistic has an  $F$ -distribution with  $n_L - 1$  numerator degrees of freedom and  $n_S - 1$  denominator degrees of freedom under the null hypothesis of equal variance and independent normal samples

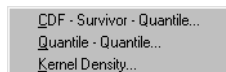
**Siegel-Tukey test.** This test statistic is reported only for tests with two subgroups ( $G = 2$ ). The test assumes the two subgroups are independent and have equal median. The test statistic is computed using the same steps as the Mann-Whitney  $U$  test for median equality described above (page 153), with a different assignment of ranks. For the Siegel-Tukey test, first rank all observations from lowest to highest. Assign rank 1 to the lowest value. Then assign rank 2 to the *highest* value and rank 3 to the second highest value. Assign rank 4 to the second lowest value and rank 5 to the third lowest value, and so on. In other words, the ranking for the Siegel-Tukey test alternates from the lowest to the highest value for every other rank. EViews reports the normal approximation to the Siegel-Tukey statistic with a continuity correction (Sheskin, 1997, pp. 196–207)).

**Bartlett test.** This test compares the logarithm of the weighted average variance with the weighted sum of the logarithms of the variances. Under the joint null hypothesis that the subgroup variances are equal and that the sample is normally distributed, the test statistic is approximately distributed as a  $\chi^2$  with  $G - 1$  degrees of freedom. Note, however, that the joint hypothesis implies that this test is sensitive to departures from normality. EViews reports the adjusted Bartlett statistic. For details, see Sokal and Rohlf (1995) and Judge, *et al.* (1985).

**Levene test.** This test is based on an analysis of variance (ANOVA) of the absolute difference from the mean. The  $F$ -statistic for the Levene test has an approximate  $F$ -distribution with  $G - 1$  numerator degrees of freedom and  $N - G$  denominator degrees of freedom under the null hypothesis of equal variances in each subgroup (Levene (1960)).

**Brown-Forsythe (modified Levene) test.** This is a modification of the Levene test in which we replace the absolute *mean* difference with the absolute *median* difference and appears to be a superior test in terms of robustness and power (Conover, *et al.* (1981), Brown and Forsythe (1974), Neter, *et al.* (1996)).

## Distribution Graphs



These views display various graphs that characterize the empirical distribution of the series. Since there are a large number of options for these views, we defer a detailed description until Chapter 9. Note, however, that the default settings for these views will generally be sufficient

for all but the most specialized analyses.

### CDF-Survivor-Quantile

This view plots the empirical cumulative distribution, survivor, and quantile functions of the series together with plus/minus two standard error bands. EViews provides a number of alternative methods for performing these computations.

## Quantile-Quantile

The quantile-quantile (QQ)-plot is a simple yet powerful tool for comparing two distributions. This view plots the quantiles of the chosen series against the quantiles of another series or a theoretical distribution.

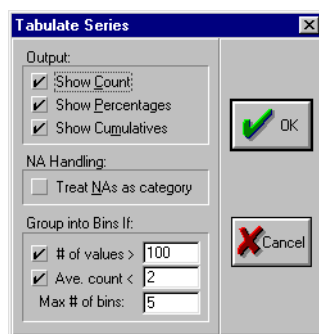
## Kernel Density

This view plots the kernel density estimate of the distribution of the series. The simplest non-parametric density estimate of a distribution of a series is the histogram. The histogram, however, is sensitive to the choice of origin and is not continuous. The kernel density estimator replaces the “boxes” in a histogram by “bumps” that are smooth (Silverman 1986). Smoothing is done by putting less weight on observations that are further from the point being evaluated.

EViews provides a number of kernel choices as well as control over bandwidth selection and computational method.

## One-Way Tabulation

This view tabulates the series in ascending order, optionally displaying the counts, percentage counts, and cumulative counts. When you select **View/One-Way Tabulation...**, the Tabulate Series dialog box appears

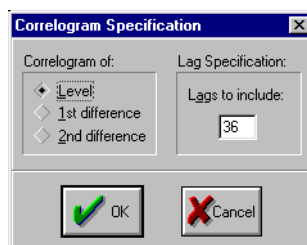


The Output options control which statistics to display in the table. You should specify the NA handling and the grouping options as described above in the discussion of “Stats by Classification” on page 145.

Cross-tabulation ( $n$ -way tabulation) is also available as a group view. See Chapter 8 for details.

## Correlogram

This view displays the autocorrelation and partial autocorrelation functions up to the specified order of lags. These functions characterize the pattern of temporal dependence in the series and typically makes sense only for time series data. When you select **View/Correlogram...**, the Correlogram Specification dialog box appears

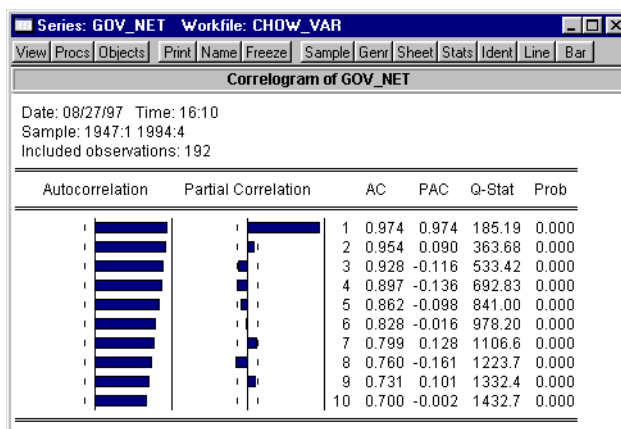


You may choose to plot the correlogram of the raw series (level)  $x$ , the first difference  $d(x) = x - x(-1)$ , or the second difference

$$d(x) - d(x(-1)) = x - 2x(-1) + x(-2)$$

of the series.

You should also specify the highest order of lag to display the correlogram; type in a positive integer in the field box. The series view displays the correlogram and associated statistics:



## Autocorrelations (AC)

The autocorrelation of a series  $y$  at lag  $k$  is estimated by:

$$\tau_k = \frac{\sum_{t=k+1}^T (y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^T (y_t - \bar{y})^2} \quad (7.15)$$

where  $\bar{y}$  is the sample mean of  $y$ . This is the correlation coefficient for values of the series  $k$  periods apart. If  $\tau_1$  is nonzero, it means that the series is first order serially correlated. If  $\tau_k$  dies

off more or less geometrically with increasing lag  $k$ , it is a sign that the series obeys a low-order autoregressive (AR) process. If  $\tau_k$  drops to zero after a small number of lags, it is a sign that the series obeys a low-order moving-average (MA) process. See Chapter 13, “Serial Correlation Theory” beginning on page 273, for a more complete description of AR and MA processes.

Note that the autocorrelations estimated by EViews differ slightly from theoretical descriptions of the estimator:

$$\tau_k = \frac{\sum_{t=k+1}^T ((y_t - \bar{y})(y_{t-k} - \bar{y})) / (T-K)}{\sum_{t=1}^T (y_t - \bar{y})^2 / T} \quad (7.16)$$

where  $\bar{y}_{t-k} = \sum y_{t-k} / (T-k)$ . The difference arises since, for computational simplicity, EViews employs the same overall sample mean  $\bar{y}$  as the mean of both  $y_t$  and  $y_{t-k}$ . While both formulations are consistent estimators, the EViews formulation biases the result toward zero in finite samples.

The dotted lines in the plots of the autocorrelations are the approximate two standard error bounds computed as  $\pm 2/\sqrt{T}$ . If the autocorrelation is within these bounds, it is not significantly different from zero at (approximately) the 5% significance level.

### Partial Autocorrelations (PAC)

The partial autocorrelation at lag  $k$  is the regression coefficient on  $y_{t-k}$  when  $y_t$  is regressed on a constant,  $y_{t-1}, \dots, y_{t-k}$ . This is a *partial* correlation since it measures the correlation of  $y$  values that are  $k$  periods apart after removing the correlation from the intervening lags. If the pattern of autocorrelation is one that can be captured by an autoregression of order less than  $k$ , then the partial autocorrelation at lag  $k$  will be close to zero.

The PAC of a pure autoregressive process of order  $p$ ,  $AR(p)$ , cuts off at lag  $p$ , while the PAC of a pure moving average (MA) process asymptotes gradually to zero.

EViews estimates the partial autocorrelation at lag  $k$  recursively by

$$\phi_k = \begin{cases} \tau_1 & \text{for } k = 1 \\ \tau_k - \frac{\sum_{j=1}^{k-1} \phi_{k-1,j} \tau_{k-j}}{1 - \sum_{j=1}^{k-1} \phi_{k-1,j} \tau_{k-j}} & \text{for } k > 1 \end{cases} \quad (7.17)$$

where  $\tau_k$  is the estimated autocorrelation at lag  $k$  and

$$\phi_{k,j} = \phi_{k-1,j} - \phi_k \phi_{k-1,k-j} \quad (7.18)$$



This is a consistent approximation of the partial autocorrelation. The algorithm is described in Box and Jenkins (1976, Part V, Description of computer programs). To obtain a more precise estimate of  $\phi$ , simply run the regression

$$y_t = \beta_0 + \beta_1 y_{t-1} + \dots + \beta_{k-1} y_{t-(k-1)} + \phi_k y_{t-k} + e_t \quad (7.19)$$

where  $e_t$  is a residual. The dotted lines in the plots of the partial autocorrelations are the approximate two standard error bounds computed as  $\pm 2/\sqrt{T}$ . If the partial autocorrelation is within these bounds, it is not significantly different from zero at (approximately) the 5% significance level.

### Q-Statistics

The last two columns reported in the correlogram are the Ljung-Box  $Q$ -statistics and their  $p$ -values. The  $Q$ -statistic at lag  $k$  is a test statistic for the null hypothesis that there is no autocorrelation up to order  $k$  and is computed as:

$$Q_{LB} = T(T+2) \sum_{j=1}^k \frac{\tau_j^2}{T-j} \quad (7.20)$$

where  $\tau_j$  is the  $j$ -th autocorrelation and  $T$  is the number of observations. If the series is not based upon the results of ARIMA estimation, then under the null hypothesis,  $Q$  is asymptotically distributed as a  $\chi^2$  with degrees of freedom equal to the number of autocorrelations. If the series represents the residuals from ARIMA estimation, the appropriate degrees of freedom should be adjusted to represent the number of autocorrelations less the number of AR and MA terms previously estimated. Note also that some care should be taken in interpreting the results of a Ljung-Box test applied to the residuals from an ARMAX specification (see Dezhbaksh, 1990, for simulation evidence on the finite sample performance of the test in this setting).

The  $Q$ -statistic is often used as a test of whether the series is white noise. There remains the practical problem of choosing the order of lag to use for the test. If you choose too small a lag, the test may not detect serial correlation at high-order lags. However, if you choose too large a lag, the test may have low power since the significant correlation at one lag may be diluted by insignificant correlations at other lags. For further discussion, see Ljung and Box (1979) or Harvey (1990, 1993).

### Unit Root Test

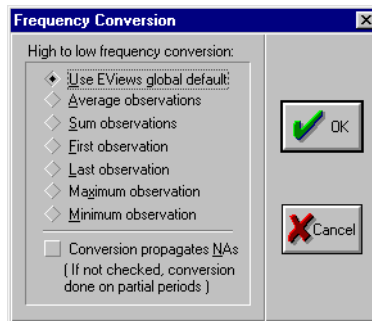
This view carries out the augmented Dickey-Fuller and Phillips-Perron unit root tests for whether the series is stationary. Select the test type, determine whether to test for a unit root in the levels, first, or second differences, whether to include an intercept and/or trend, and the number of lags to include in the test regression.

See Chapter 13 for a discussion of stationary and nonstationary time series and additional details on how to carry out the unit root tests in EViews (page 297).

## Conversion Options

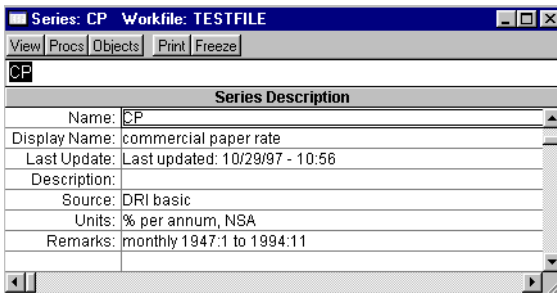
When you fetch a series from an EViews database, it will automatically be converted to the frequency of the current workfile. The conversion options view allows you to set the method that will be used to perform these conversions.

Each series has a frequency conversion method associated with it. The default conversion method is the one set by the EViews global options; see Appendix B. You can change or check the conversion method by clicking **View/Conversion Options...**



## Label

This view displays the description of the series, which might look as follows:



You can edit any of the field cells in the series label, except the **Last Update** cell which displays the date/time the series was last modified. Each field contains a single line, except for the **Remarks and History** fields which can contain up to 20 comment lines. Note that if you insert a line, the last (of the 20) line of these fields will be deleted.

The **Name** is the series name as it appears in the workfile; you can rename your series by editing this cell. If you fill in the **Display Name** field, this name may be used in tables and graphs in place of the standard object name. Unlike ordinary object names, **Display Names** may contain spaces and preserve capitalization (upper and lower case letters).

See Chapter 6 for further discussion of label fields and their use in Database searches.

## Series Procedures

Generate by Equation...  
 Seasonal Adjustment...  
 Exponential Smoothing...  
 Hodrick-Prescott Filter...

Series procedures may be used to generate new series that are based upon the data in the original series.

## Generate by Equation

This is a general procedure that allows you to create new series by using expressions to transform values in existing series. The rules governing the generation of series are explained in detail in Chapter 5.

## Seasonal Adjustment

Time series observed at quarterly and monthly frequencies often exhibit cyclical movements that recur every month or quarter. For example, ice cream sales may surge during summer every year and toy sales may reach a peak every December during Christmas sales. Seasonal adjustment refers to the process of removing these cyclical seasonal movements from a series and extracting the underlying trend component of the series.

Seasonal adjustment procedures are available only for quarterly and monthly series. To seasonally adjust a series, click on **Procs/Seasonal Adjustment...** in the series window toolbar. The Seasonal Adjustment dialog box will appear.

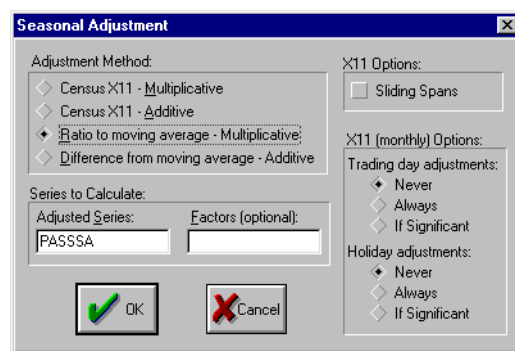
You must provide the following information:

**Adjustment Method.** Census X-11 is the standard U.S. Bureau of the Census adjustment method. The other two methods are less sophisticated approaches to seasonal adjustment.

There are two classes of seasonal adjustment methods, multiplicative and additive. Multiplicative methods are appropriate if the series can be decomposed into a product of a trend and a seasonal component, while additive methods are used if the series can be decomposed into a sum of a trend and a seasonal component. Multiplicative methods (either the X-11 or ratio-to-moving average option) should only be used with series that are always positive.

**Adjusted Series.** This is the name for the seasonally adjusted series. EViews proposes a name by appending SA to the name of the original series, but you can change it if you want. You may optionally provide a name for the seasonal factor series, in which case it will be saved in the workfile.

**Factors.** You may optionally provide a name for the seasonal to saved the series in the workfile. For moving average methods, out-of-sample values will be filled in with in-sample estimates of the seasonal factors. For the X11 methods, future values of the seasonal factors will



be filled in with one-year-ahead forecasts of the seasonal factors (for values more than one year ahead, we simply repeat the one year ahead forecasts).

### Census X-11 Methods

The Census X-11 methods (multiplicative and additive) are the standard methods used by the U.S. Bureau of Census to seasonally adjust publicly released data. The X-11 routines are separate programs provided by the Census and are installed in the EViews directory in the files X11Q2.EXE and X11SS.EXE. The documentation for these programs can also be found in your EViews directory as text files X11DOC1.TXT through X11DOC3.TXT.

The X-11 programs may be run directly from DOS or from inside EViews. If you run the X-11 programs from EViews by choosing the X-11 options in the Seasonal Adjustment dialog, the adjusted series and the factor series will be automatically imported into your EViews workfile. X-11 summary output and error messages will also be displayed in the series window at the end of the procedure.

The X-11 method has many options, the most important of which are available in the Seasonal Adjustment dialog. However, there are other options not available in the EViews dialog; to use these other options, you should run the X-11 programs from the DOS command line. All options available in the X-11 methods are described in the X11DOC text files in your EViews directory.

You should note that there is a limit on the number of observations that you can seasonally adjust. X-11 only works for quarterly and monthly frequencies, requires at least four full years of data, and can adjust only up to 20 years of monthly data and up to 30 years of quarterly data.

### Ratio to moving average—multiplicative

The algorithm works as follows. Denote the series to be filtered by  $y_t$ .

1. First compute the centered moving average of  $y_t$  as

$$x_t = \begin{cases} (0.5y_{t+6} + \dots + y_t + \dots + 0.5y_{t-6})/12 & \text{for monthly series} \\ (0.5y_{t+2} + y_{t+1} + y_t + y_{t-1} + 0.5y_{t-2})/4 & \text{for quarterly series} \end{cases} \quad (7.21)$$

2. Take the ratio  $\tau_t = y_t/x_t$ .
3. Compute the seasonal indices. For monthly series, the seasonal index  $i_m$  for month  $m$  is the average of  $\tau_t$  using observations only for month  $m$ . For quarterly series, the seasonal index  $i_q$  for quarter  $q$  is the average of  $\tau_t$  using observations only for quarter  $q$ .
4. We then adjust the seasonal indices so that they multiply to one. This is done by computing the seasonal factors as the ratio of the seasonal index to the geometric mean of the indices:

$$s = \begin{cases} i_m / (\sqrt[12]{i_1 i_2 \dots i_{12}}) & \text{for monthly series} \\ i_q / (\sqrt[4]{i_1 i_2 i_3 i_4}) & \text{for quarterly series} \end{cases} \quad (7.22)$$

5. These  $s$  are the reported *scaling factors* in the series window and are saved as series if you provide a name in the field box. The interpretation is that the series  $y$  is  $s_j$  percent higher in period  $j$  relative to the adjusted series.
6. The seasonally adjusted series is obtained by dividing  $y_t$  by the seasonal factors  $s_j$ .

### Difference from moving average—additive

Suppose that we wish to filter  $y_t$ .

1. First compute the centered moving average of  $y_t$  as in Equation (7.21) on page 162.
2. Take the difference  $d_t = y_t - x_t$ .
3. Compute the seasonal indices. For monthly series, the seasonal index  $i_m$  for month  $m$  is the average of  $d_t$  using observations only for month  $m$ . For quarterly series, the seasonal index  $i_q$  for quarter  $q$  is the average of  $d_t$  using observations only for quarter  $q$ .
4. We then adjust the seasonal indices so that they add up to zero. This is done by setting  $s_j = i_j - \bar{i}$  where  $\bar{i}$  is the average of all seasonal indices. These  $s$  are the reported *scaling factors*. The interpretation is that the series  $y$  is  $s_j$  higher in period  $j$  relative to the adjusted series.
5. The seasonally adjusted series is obtained by subtracting the seasonal factors  $s_j$  from  $y_t$ .

The main difference between X-11 and the moving average methods is that the seasonal factors may change from year to year in X-11. The seasonal factors are assumed to be constant for the moving average method.

### Exponential Smoothing

Exponential smoothing is a simple method of adaptive forecasting. It is an effective way of forecasting when you have only a few observations on which to base your forecast. Unlike forecasts from regression models which use fixed coefficients, forecasts from exponential smoothing methods adjust based upon past forecast errors. Below, we provide a brief discussion of exponential smoothing methods in EViews. For additional discussion, see Bowerman and O'Connell (1979).

To obtain forecasts based on exponential smoothing methods, choose **Procs/Exponential Smoothing**. The Exponential Smoothing dialog box appears:

**Exponential Smoothing**

Smoothing Method: # Parameters

- Single 1
- Double 1
- Holt-Winters - No seasonal 2**
- Holt-Winters - Additive 3
- Holt-Winters - Multiplicative 3

Smoothing Parameters:

Alpha: (mean)  Enter number between 0 and 1, or E to estimate.

Beta: (trend)

Gamma: (seasonal)

Smoothed Series:   
Series name for smoothed and forecasted values.

Estimation Sample:   
Forecasts begin in period following estimation endpoint.

Cycle for Seasonal:

You need to provide the following information:

**Smoothing Method.** You have the option to choose from five methods, each of which are described below.

**Smoothing Parameters.** You can either specify the values of the smoothing parameters or let EViews estimate them.

To estimate the parameter, type the letter  $e$  (for estimate) in the edit field. EViews estimates the parameters by minimizing the sum of squared errors. Don't be surprised if the estimated damping parameters are close to one—it is a sign that the series is close to a random walk, where the most recent value is the best estimate of future values.

To specify a number, type the number in the field corresponding to the parameter. All parameters are constrained to be between 0 and 1; if you specify a number outside the

unit interval, EViews will estimate the parameter.

**Smoothed Series Name.** You should provide a name for the smoothed series. EViews suggests a name by appending SM to the original series name but you can change the name to your liking.

**Estimation Sample.** You must specify the sample period upon which to base your forecasts (whether or not you choose to estimate the parameters). The default is the current workfile sample. EViews will calculate forecasts starting from the first observation after the end of the estimation sample.

**Cycle for Seasonal.** You can change the number of seasons per year from the default of 12 for monthly or 4 for quarterly series. This option allows you to forecast from unusual data such as an undated workfile. Enter a number for the cycle in this field.

### Single Smoothing (one parameter)

This single exponential smoothing method is appropriate for series that move randomly above and below a constant mean with no trend nor seasonal patterns. The smoothed series  $\hat{y}_t$  of  $y_t$  is computed recursively by evaluating

$$\hat{y}_t = \alpha y_t + (1 - \alpha) \hat{y}_{t-1} \quad (7.23)$$

where  $0 < \alpha \leq 1$  is the *damping* (or *smoothing*) factor. The smaller is the  $\alpha$ , the smoother is the  $\hat{y}_t$  series. By repeated substitution, we can rewrite the recursion as

$$\hat{y}_t = \alpha \sum_{s=0}^{t-1} (1 - \alpha)^s y_{t-s} \quad (7.24)$$

This shows why this method is called exponential smoothing—the forecast of  $y_t$  is a weighted average of the past values of  $y_t$ , where the weights decline exponentially with time.

The forecasts from single smoothing are constant for all future observations. This constant is given by

$$\hat{y}_{T+k} = \hat{y}_T \quad \text{for all } k > 0 \quad (7.25)$$

where  $T$  is the end of the estimation sample.

To start the recursion, we need an initial value for  $\hat{y}_t$  and a value for  $\alpha$ . EViews uses the mean of the initial observations of  $y_t$  to start the recursion. Bowerman and O'Connell (1979) suggest that values of  $\alpha$  around .01 to .30 work quite well. You can also let EViews estimate  $\alpha$  to minimize the sum of squares of one-step forecast errors.

### Double smoothing (one parameter)

This method applies the single smoothing method twice (using the same parameter) and is appropriate for series with a linear trend. Double smoothing of a series  $y$  is defined by the recursions

$$\begin{aligned} S_t &= \alpha y_t + (1 - \alpha)S_{t-1} \\ D_t &= \alpha S_t + (1 - \alpha)D_{t-1} \end{aligned} \quad (7.26)$$

where  $S$  is the single smoothed series and  $D$  is the double smoothed series. Note that double smoothing is a single parameter smoothing method with damping factor  $0 < \alpha \leq 1$ .

Forecasts from double smoothing are computed as

$$\begin{aligned} \hat{y}_{T+k} &= \left(2 + \frac{\alpha k}{1 - \alpha}\right)S_T - \left(1 + \frac{\alpha k}{1 - \alpha}\right)D_T \\ &= \left(2S_T - D_T + \frac{\alpha}{1 - \alpha}(S_T - D_T)k\right) \end{aligned} \quad (7.27)$$

The last expression shows that forecasts from double smoothing lie on a linear trend with intercept  $2S_T - D_T$  and slope  $\alpha(S_T - D_T)/(1 - \alpha)$ .

### Holt-Winters—multiplicative (three parameters)

This method is appropriate for series with a linear time trend and multiplicative seasonal variation. The smoothed series  $\hat{y}_t$  is given by

$$\hat{y}_{t+k} = (a + bk)c_{t+k} \quad (7.28)$$

where

$$\begin{array}{ll}
a & \text{permanent component (intercept)} \\
b & \text{trend} \\
c_t & \text{multiplicative seasonal factor}
\end{array}
\tag{7.29}$$

These three coefficients are defined by the following recursions

$$\begin{aligned}
a(t) &= \alpha \frac{y_t}{c_t(t-s)} + (1-\alpha)(a(t-1) + b(t-1)) \\
b(t) &= \beta(a(t) - a(t-1)) + (1-\beta)b(t-1) \\
c_t(t) &= \gamma \frac{y_t}{a(t)} + (1-\gamma)c_t(t-s)
\end{aligned}
\tag{7.30}$$

where  $0 < \alpha, \beta, \gamma < 1$  are the damping factors and  $s$  is the seasonal frequency specified in the **Cycle for Seasonal** field box.

Forecasts are computed by

$$\hat{y}_{t+k} = (a(T) + b(T)k)c_{T+k-s} \tag{7.31}$$

where the seasonal factors are used from the last  $s$  estimates.

#### Holt-Winters—additive (three parameter)

This method is appropriate for series with a linear time trend and additive seasonal variation. The smoothed series  $\hat{y}_t$  is given by

$$\hat{y}_{t+k} = a + bk + c_{t+k} \tag{7.32}$$

where  $a, b$  are the permanent component and trend as defined above in Equation (7.29) and  $c_t$  is the additive seasonal factor. The three coefficients are defined by the following recursions

$$\begin{aligned}
a(t) &= \alpha(y_t - c_t(t-s)) + (1-\alpha)(a(t-1) + b(t-1)) \\
b(t) &= \beta(a(t) - a(t-1)) + (1-\beta)b(t-1) \\
c_t(t) &= \gamma(y_t - a(t+1)) - \gamma c_t(t-s)
\end{aligned}
\tag{7.33}$$

where  $0 < \alpha, \beta, \gamma < 1$  are the damping factors and  $s$  is the seasonal frequency specified in the **Cycle for Seasonal** field box.

Forecasts are computed by

$$\hat{y}_{T+k} = a(T) + b(T)k + c_{T+k-s} \tag{7.34}$$

where the seasonal factors are used from the last  $s$  estimates.



### Holt-Winters—no seasonal (two parameters)

This method is appropriate for series with a linear time trend and no seasonal variation. This method is similar to the double smoothing method in that both generate forecasts with a linear trend and no seasonal component. The double smoothing method is more parsimonious since it uses only one parameter, while this method is a two parameter method. The smoothed series  $\hat{y}_t$  is given by

$$\hat{y}_{t+k} = a + bk \quad (7.35)$$

where  $a, b$  are the permanent component and trend as defined above in Equation (7.29).

These two coefficients are defined by the following recursions

$$\begin{aligned} a(t) &= \alpha y_t + (1 - \alpha)(a(t-1) + b(t-1)) \\ b(t) &= \beta(a(t) - a(t-1)) + 1 - \beta b(t-1) \end{aligned} \quad (7.36)$$

where  $0 < \alpha, \beta < 1$  are the damping factors. This is an exponential smoothing method with two parameters.

Forecasts are computed by

$$\hat{y}_{T+k} = a(T) + b(T)k \quad (7.37)$$

These forecasts lie on a linear trend with intercept  $a(T)$  and slope  $b(T)$ .

It is worth noting that Holt-Winters—no seasonal is *not* the same as additive or multiplicative with  $\gamma = 0$ . The condition  $\gamma = 0$  only restricts the seasonal factors from changing over time so there are still (fixed) nonzero seasonal factors in the forecasts.

### Illustration

As an illustration of forecasting using exponential smoothing we forecast data on monthly housing starts (HS) for the period 1985:01–1988:12 using the DRI Basics data for the period 1959:01–1984:12. These data are provided in the workfile HS.WF1. Load the workfile, highlight the HS series, double click, select **Procs/Exponential Smoothing...** We use the Holt-Winters—multiplicative method to account for seasonality, name the smoothed forecasts as HS\_SM, and estimate all parameters over the period 1959.1–1984.12.

When you click **OK**, EViews displays the results of the smoothing procedure. The first part displays the estimated (or specified) parameter values, the sum of squared residuals, the root mean squared error of the forecast. The zero values for beta and gamma in this example means that the trend and seasonal components are estimated as fixed and not changing.

Date: 10/15/97 Time: 00:57  
Sample: 1959:01 1984:12  
Included observations: 312  
Method: Holt-Winters Multiplicative Seasonal  
Original Series: HS  
Forecast Series: HS SM

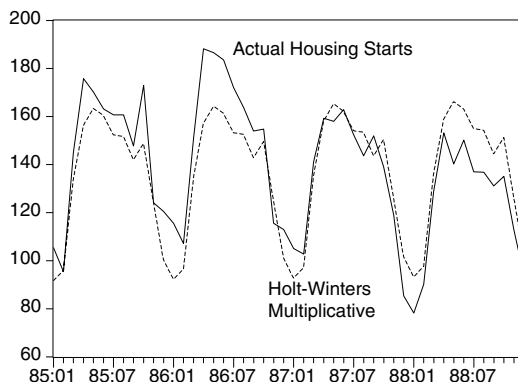
Parameters:	Alpha	0.7100
	Beta	0.0000
	Gamma	0.0000
Sum of Squared Residuals		40365.69
Root Mean Squared Error		11.37441

The second part of the table displays the mean ( $\alpha$ ) and trend ( $\beta$ ) at the end of the estimation sample that are used for post-sample smoothed forecasts.

End of Period Levels:	Mean	134.6584
	Trend	0.064556
Seasonals:	1984:01	0.680745
	1984:02	0.711559
	1984:03	0.992958
	1984:04	1.158501
	1984:05	1.210279
	1984:06	1.187010
	1984:07	1.127546
	1984:08	1.121792
	1984:09	1.050131
	1984:10	1.099288
	1984:11	0.918354
	1984:12	0.741837

For seasonal methods, the seasonal factors ( $\gamma$ ) used in the forecasts are also displayed. The smoothed series in the workfile contains data from the beginning of the estimation sample to the end of the workfile range; all values after the estimation period are forecasts.

When we plot the actual values and the smoothed forecasts on a single graph, we get:



The forecasts from the multiplicative exponential smoothing method do a good job of tracking the seasonal movements in the actual series.

### Hodrick-Prescott Filter

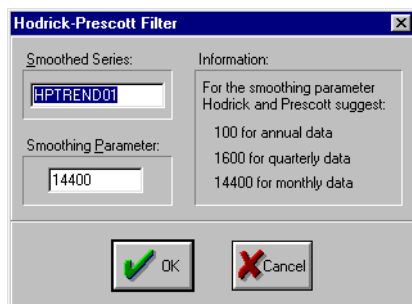
This is a smoothing method that is widely used among macroeconomists to obtain a smooth estimate of the long-term trend component of a series. The method was first used in a working paper (circulated in the early 1980's and published in 1997) by Hodrick and Prescott to analyze postwar U.S. business cycles.

Technically, the Hodrick-Prescott (HP) filter is a two-sided linear filter that computes the smoothed series  $s$  of  $y$  by minimizing the variance of  $y$  around  $s$ , subject to a penalty that constrains the second difference of  $s$ . That is, the HP filter chooses  $s_t$  to minimize:

$$\sum_{t=1}^T (y_t - s_t)^2 + \lambda \sum_{t=2}^{T-1} ((s_{t+1} - s_t) - (s_t - s_{t-1}))^2. \quad (7.38)$$

The penalty parameter  $\lambda$  controls the smoothness of the series  $s_t$ . The larger the  $\lambda$ , the smoother the  $s_t$ . As  $\lambda \rightarrow \infty$ ,  $s_t$  approaches a linear trend.

To smooth the series using the Hodrick-Prescott filter, choose **Procs/Hodrick-Prescott Filter**.... The Hodrick-Prescott Filter dialog box appears:



First, provide a name for the smoothed series. EViews will suggest a name, but you can always enter a name of your choosing. Next, specify an integer value for the smoothing parameter,  $\lambda$ . The default values in EViews are set to be:

$$\lambda = \begin{cases} 100 & \text{for annual data} \\ 1,600 & \text{for quarterly data} \\ 14,400 & \text{for monthly data} \end{cases} \quad (7.39)$$

EViews will round off any non-integer values.

When you click **OK**, EViews displays a graph of the filtered series together with the original series. Note that only data in the current workfile sample are filtered. Data for the smoothed series outside the current sample are filled with NAs.

## Commands

The command syntax is to follow the name of the series with a dot and then the view or procedure name, with options specified in parentheses. For example, to view the histogram and descriptive statistics of a series named LWAGE, type

```
lwage.hist
```

To test whether the mean of the series HRS is equal to 3, type

```
hrs.teststat(mean=3)
```

To plot the quantiles of the series INJURE against the quantiles of the uniform distribution, type

```
injure.qqplot(u)
```

To plot the correlogram of the series GDP up to 20 lags, type

```
gdp.correl(20)
```

To smooth the series GDP by the Hodrick-Prescott filter with smoothing parameter 1600 and to save the smoothed series as GDP\_HP, type

```
gdp.hp(1600) gdp_hp
```

See the *Command and Programming Reference* for a complete list of commands and options that are available for series objects.



## Chapter 8. Groups

---

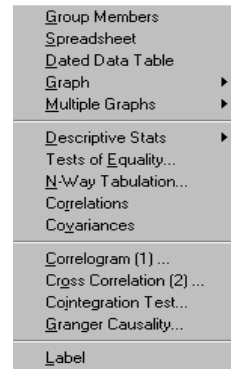
This chapter describes the views and procedures of a group object. With a group, you can compute various statistics that describe the relationship between multiple series and display them in various forms such as spreadsheets, tables, and graphs.

The remainder of this chapter assumes that you are already familiar with the basics of creating and working with a group. Chapters 3—6 provide relevant details on the basic operations.

### Views from a Group Window

The group view drop-down menu is divided into four blocks:

- The views in the first block provide various ways of looking at the actual data in the group.
- The views in the second block display various basic statistics.
- The views in the third block are for specialized statistics for time series data.
- The fourth block contains the label view, which provides information regarding the group object.



### Group Members

This view displays the member series in the group and allows you to alter the group. To change the group, simply edit the group window. You can add other series from the workfile, include expressions involving series, or you can delete series from the group.

Note that editing the window does not change the list of group members. Once you make your changes to the list, you must press the **UpdateGroup** button in the group window toolbar to save the changes.

### Spreadsheet

This view displays the data, in spreadsheet form, for each series in the group.

If you wish, you can flip the rows and columns of the spreadsheet by pressing the **Transpose** button. In transpose format, each row contains a series, and each column

an observation or date. Pressing the **Transpose** button toggles between the two spreadsheet views.

Dated Data Table

The dated data table view is used to construct tables for reporting and presenting data, forecasts, and simulation results. This view displays the series contained in the group in a variety of formats. You can also use this view to perform common transformations and frequency conversions, and to display data at various frequencies in the same table.

For example, suppose you wish to show your quarterly data for the CS and GDP series, with data for each year, along with an annual average, on a separate line:

	1988				1988
CS	3128.2	3147.8	3170.6	3202.9	3162.4
GDP	4655.3	4704.8	4734.5	4779.7	4718.6
	1989				1989
CS	3203.6	3212.2	3235.3	3242.0	3223.3
GDP	4817.6	4839.0	4839.0	4856.7	4838.1

The dated data table handles all of the work of setting up this table, and computing the summary values.

Alternatively, you may wish to display annual averages for each year up to the last, followed by the last four quarterly observations in your sample:

	1987	1988	1989	89:1	89:2	89:3	89:4
CS	3052.2	3162.4	3223.3	3203.6	3212.2	3235.3	3242.0
GDP	4540.0	4718.6	4838.1	4817.6	4839.0	4839.0	4856.7

Again, the dated data table may be used to automatically perform the required calculations and set up the table.

The dated data table is capable of creating more complex tables and performing a variety of other calculations. Note, however, that the dated data table view is currently available only for annual, semi-annual, quarterly, or monthly workfiles.

Creating and Specifying a Dated Data Table

To create a dated data table, create a group containing the series of interest and select **View/ Dated Data Table**. The group window initially displays a default table view. The default is to show a single year of data on each line, along with a summary measure (the annual average).



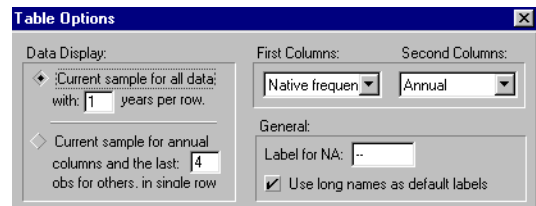
You can, however, set options to control the display of your data through the Table and Row Options dialogs. Note the presence of two new buttons on the group window toolbar, labeled **TabOptions** (for Table Options) and **RowOptions**. **TabOptions** sets the global options for the dated data table. These options will apply to all series in the group object. The **RowOptions** button allows you to override the global options for a particular series. Once you specify table and row options for your group, EViews will remember these options the next time you open the dated data table view for the group.

### Table Setup

When you click on the **TabOptions** button, the Table Options dialog appears. The top half of the dialog provides options to control the general style of the table.

The buttons on the left hand side of the dialog allow you to choose between the two display formats described above:

- The first style displays the data for  $n$  years per row, where  $n$  is the positive integer specified in the edit field.
- The second style is a bit more complex. It allows you to specify, for data displayed at a frequency other than annual, the number of observations taken *from the end of the workfile sample* that are to be displayed. For data displayed at an annual frequency, EViews will display observations over the entire workfile sample.



The two combo boxes on the top right of the dialog supplement your dated display choice by allowing you to display your data at multiple frequencies in each row. The **First Columns** selection describes the display frequency for the first group of columns, while the **Second Columns** selection controls the display for the second group of columns. If you select the same frequency, only one set of results will be displayed.

In each combo box, you may choose among:

- Native frequency (the frequency of the workfile)
- Annual
- Quarterly
- Monthly

If necessary, EViews will perform any frequency conversion (to a lower frequency) required to construct the table.

The effects of these choices on the table display are best described by the following example. For purposes of illustration, assume that the current workfile sample is quarterly, with a range from 1997:1–2003:4.

Now suppose that you choose to display the first style (two years per row), with the first columns set to the native frequency, and the second columns set to annual frequency. Each row will contain eight quarters of data (the native frequency data) followed by the corresponding two annual observations (the annual frequency data):

	1997				1998				1997	1998
GDPR	1775.4	1784.9	1789.8	1799.6	1810.6	1820.9	1830.9	1841.0	1787.4	1825.8
CS	681.0	687.2	691.6	696.0	700.2	704.4	708.7	713.3	688.9	706.6
PCGDPD	2.42	1.44	1.46	1.43	1.54	1.64	1.73	1.81	1.69	1.68
	1999				2000				1999	2000
GDPR	1851.0	1860.9	1870.7	1880.9	1891.0	1900.9	1910.6	1920.4	1865.9	1905.7
CS	717.8	722.3	726.8	731.8	736.4	741.0	745.7	750.7	724.7	743.5
PCGDPD	1.88	1.94	2.00	2.05	2.09	2.13	2.16	2.20	1.97	2.15

EViews automatically performs the frequency conversion to annual data using the specified method (see page We emphasize that the above transformation methods represent only the most commonly employed transformations. If you wish to construct your table with other transformations, you should add an appropriate auto-series to the group.).

If you reverse the ordering of data types in the first and second columns so that the first columns display the annual data, and the second columns display the native frequency, the top of the table will contain:

	1997	1998	1997				1998			
GDPR	1787.4	1825.8	1775.4	1784.9	1789.8	1799.6	1810.6	1820.9	1830.9	1841.0
CS	688.9	706.6	681.0	687.2	691.6	696.0	700.2	704.4	708.7	713.3
PCGDPD	1.69	1.68	2.42	1.44	1.46	1.43	1.54	1.64	1.73	1.81
	1999	2000	1999				2000			
GDPR	1865.9	1905.7	1851.0	1860.9	1870.7	1880.9	1891.0	1900.9	1910.6	1920.4
CS	724.7	743.5	717.8	722.3	726.8	731.8	736.4	741.0	745.7	750.7
PCGDPD	1.97	2.15	1.88	1.94	2.00	2.05	2.09	2.13	2.16	2.20

Now, click on **TabOptions**, choose the second display style, and enter 4 in the edit box. Then specify **Annual frequency** for the first columns and **Native frequency** for the second columns. EViews will display the annual data for the workfile sample, followed by the last four quarterly observations:

	1997	1998	1999	2000	2001	2002	2003	03:1	03:2	03:3	03:4
GDPR	1787.4	1825.8	1865.9	1905.7	1945.4	1985.3	2024.2	2009.8	2019.4	2029.0	2038.5
CS	688.9	706.6	724.7	743.5	763.4	783.9	805.0	797.0	802.3	807.6	813.0
PCGDPD	1.69	1.68	1.97	2.15	2.26	2.33	2.37	2.36	2.36	2.37	2.38

The last four columns display the data for the last four observations of the current sample: 2003:1–2003:4. The remaining columns are the annual transformations of the current workfile sample 1997:1–2003:4.

### Additional Table Options

The bottom of the Table Options dialog controls the default data transformations and numeric display for each series in the group. EViews allows you to use two rows, each with a different transformation and a different output format, to describe each series.

For each row, you specify the transformation method, frequency conversion method, and the number format.

Keep in mind that you may override the default transformation for a particular series using the **RowOptions** menu (see page Row Options).

### Transformation Methods

The following transformations are available:

<b>None (raw data)</b>	No transformation
<b>1 Period Difference</b>	$y - y(-1)$
<b>1 Year Difference</b>	$y - y(-f), \text{ where } f = \begin{cases} 1 & \text{for annual} \\ 2 & \text{for semi-annual} \\ 4 & \text{for quarterly} \\ 12 & \text{for monthly} \end{cases}$
<b>1 Period % Change</b>	$100 \times (y - y(-1)) / y(-1)$
<b>1 Period % Change at Annual Rate</b>	Computes $R$ such that $(1 + r/100)^f$ where $f = 1 + R/100$ and $r$ is the 1 period % change.

<b>1 Year % Change</b>	$100 \times (y - y(-f)) / y(-f)$ , where $f$ is defined above.
<b>No second row</b>	Do not display a second row

We emphasize that the above transformation methods represent only the most commonly employed transformations. If you wish to construct your table with other transformations, you should add an appropriate auto-series to the group.

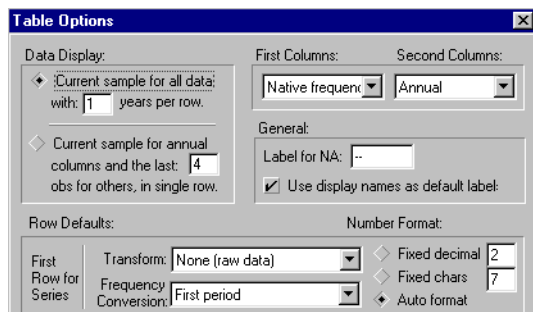
## Frequency Conversion

The following frequency conversion methods are provided:

<b>Average then Transform</b>	First convert by taking the average, then transform the average, as specified.
<b>Transform then Average</b>	First transform the series, then take the average of the transformed series.
<b>Sum then Transform</b>	First convert by taking the sum, then transform the sum, as specified.
<b>First Period</b>	Convert by taking the first quarter of each year or first month of each quarter/year.
<b>Last Period</b>	Convert by taking the last quarter of each year or last month of each quarter/year.

The choice between **Average then Transform** and **Transform then Average** changes the ordering of the transformation and frequency conversion operations. The methods differ only for non-linear transformations (such as the % change).

For example, if we specify:



EViews will display a table with data formatted in the following fashion:

	1997				1997
GDP	1775.4	1784.9	1789.8	1799.6	7149.8
CS	681.0	687.2	691.6	696.0	2755.7
	1998				1998
GDP	1810.6	1820.9	1830.9	1841.0	7303.3
CS	700.2	704.4	708.7	713.3	2826.5

If, instead, you change the **Frequency Conversion** to **First period**, EViews will display a table of the form:

	1997				1997
GDP	1775.4	1784.9	1789.8	1799.6	1775.4
CS	681.0	687.2	691.6	696.0	681.0
	1998				1998
GDP	1810.6	1820.9	1830.9	1841.0	1810.6
CS	700.2	704.4	708.7	713.3	700.2

Below, we provide an example which illustrates the computation of the percentage change measures.

### Formatting Options

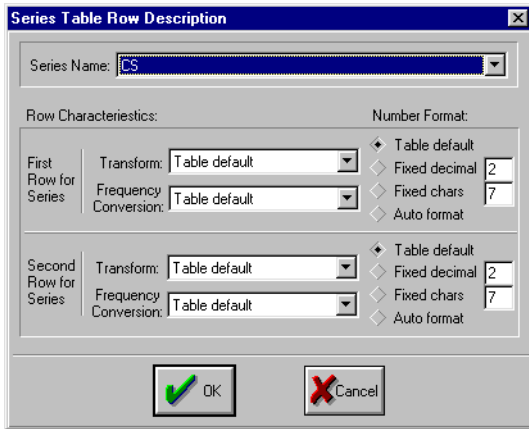
EViews lets you choose between fixed decimal, fixed digit, and auto formatting of the numeric data. Generally, auto formatting will produce attractive output, but if not, simply select the desired method and enter an integer in the appropriate box. The options are:

<b>Auto format</b>	EViews chooses the format depending on the data.
<b>Fixed decimal</b>	Specify how many digits to display after the decimal point. This option aligns all numbers at the decimal point.
<b>Fixed chars</b>	Specify how many total characters to display for each number.

EViews will round your data prior to display in order to fit the specified format. This rounding is for display purposes only and does not alter the original data.

### Row Options

These options allow you to override the row defaults specified by the Table Options dialog. You can specify a different transformation, frequency conversion method, and number format, for each series.



In the dialog that appears, select the series for which you wish to override the table default options. Then specify the transformation, frequency conversion, or number format you want to use for that series. The options are the same as those described above for the row defaults.

### Other Options

**Label for NA:** allows you to define the symbol used to identify missing values in the table. Bear in mind that if you choose to display your data in transformed form, the transformation may generate missing values even if none of the raw data are missing. Dated data table transformations are

explained above.

If your series has display names, you can use the display name as the label for the series in the table by selecting the **Use display names as default labels** option. See Chapter 3 for a discussion of display names and the label view.

### Illustration

As an example, consider the following dated data table which displays both quarterly and annual data for GDPR and CS:

	1997				1997
GDPR	1775.4	1784.9	1789.8	1799.6	1787.4
	1.20	0.54	0.27	0.55	3.19
CS	681.0	687.2	691.6	696.0	688.9
	0.95	0.91	0.63	0.64	3.17
	1998				1998
GDPR	1810.6	1820.9	1830.9	1841.0	1825.8
	0.61	0.57	0.55	0.55	2.15
CS	700.2	704.4	708.7	713.3	706.6
	0.61	0.60	0.60	0.65	2.57

The first row of output for each of the series is not transformed, while the second row contains the 1-period percentage change in the series. The second row of GDPR is computed using the **Average then Transformed** setting while the second row of CS is computed with the **Transform then Average** option. We specified this alternative setting for CS by using the **RowOptions** dialog to override the table defaults.

The first four columns show the data in native frequency so the choice between **Average then Transform** and **Transform then Average** is irrelevant—each entry in the second row measures the 1-period (1-quarter) percentage change in the variable.

The 1-period percentage change in the last column is computed differently under the two methods. The **Average then Transformed** percentage change in GDP for 1998 measures the percentage change between the average value in 1997 and the average value in 1998. It is computed as:

$$100 \cdot (1825 - 1787) / 1787 \cong 2.15 \quad (8.1)$$

EViews computes this transformation using full precision for intermediate results, then displays the result using the specified number format.

The computation of the **Transform then Average** one-period change in CS for 1998 is a bit more subtle. Since we want a measure of the annual change, we first evaluate the *one-year* percentage change at each of the quarters in the year, and then average the results. For example, the one-year percentage change in 1998:1 is given by  $100 \cdot (700.2 - 681.0) / 681.0 \cong 2.82$  and the one-year percentage change in 1998:2 is  $100 \cdot (704.4 - 687.2) / 687.2 \cong 2.50$ . Averaging these percentage changes yields

$$100 \cdot \frac{1}{4} \left( \frac{700.2 - 681.0}{681.0} + \frac{704.5 - 687.2}{687.2} + \frac{708.7 - 691.6}{691.6} + \frac{713.3 - 696.0}{696.0} \right) \cong 2.57$$

Note that this computation differs from evaluating the average of the one-quarter percentage changes for each of the quarters of the year.

## Other Menu Items

**Edit + /-** allows you to edit the row (series) labels as well as the actual data in the table. You will not be able to edit any of the computed ranks and any changes that you make to the row labels will only apply to the dated data table view.

We warn you that *if you edit a data cell, the underlying series data will also change*. This latter feature allows you to use dated data tables for data entry from published sources.

If you want to edit the data in the table but wish to keep the underlying data unchanged, first **Freeze** the table view and then apply **Edit** to the frozen table.

**Font** allows you to choose the font, font style, and font size to be used in the table.

**Title** allows you to add a title to the table.

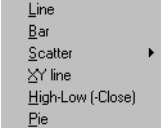
**Sample** allows you to change the sample to display in the table.

Here is an example of a table after freezing and editing:

	1997				1997
Gross Domestic Product (Billions \$ '92)	1775.4	1784.9	1789.8	1799.6	1787.4
One-period % change	(1.20)	(0.54)	(0.27)	(0.55)	(3.19)
Consumer Expenditure - Services (Billions \$ '92)	681.0	687.2	691.6	696.0	688.9
One-period % change	(0.95)	(0.91)	(0.63)	(0.64)	(3.17)
	1998				1998
Gross Domestic Product (Billions \$ '92)	1810.6	1820.9	1830.9	1841.0	1825.8
One-period % change	(0.61)	(0.57)	(0.55)	(0.55)	(2.15)
Consumer Expenditure - Services (Billions \$ '92)	700.2	704.4	708.7	713.3	706.6
One-period % change	(0.61)	(0.60)	(0.60)	(0.65)	(2.57)

## Graphs

These views display the series in the group in various graphical forms. You can create graph objects by freezing these views. Chapter 10, “Graphs, Tables, and Text Objects” explains how to edit and modify graph objects in EViews.



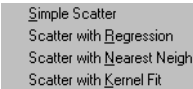
The **Graph** views display all series in a single graph. To display each series in a separate graph, see **Multiple Graphs**, page 184 below.

### Line and Bar Graphs

Displays a line or bar graph of the series in the group. Click anywhere in the background of the graph to modify the scaling options or line patterns.

### Scatter

There are four variations on the scatter diagram view of a series.



**Simple Scatter** plots a scatter diagram with the first series on the horizontal axis and the remaining series on the vertical axis.

The remaining three options, **Scatter with Regression**, **Scatter with Nearest Neighbor Fit**, and **Scatter with Kernel Fit** plot fitted lines of the first series against the second on top of the scatter diagram. They differ in how the fitted lines are calculated. All three graph views are described in detail in Chapter 9 (“Scatter Diagrams with Fit Lines”, beginning on page 205).

### XY Line

This view plots the first series on the horizontal X-axis and the remaining series on the vertical Y-axis. XY Line is similar to Simple Scatter with two substantive differences. First, XY Lines are displayed in a rectangle (1:1-1/3 ratio) while scatterplots are displayed in a square (1: 1) box. Second, points are connected in XY Lines but not in scatters.

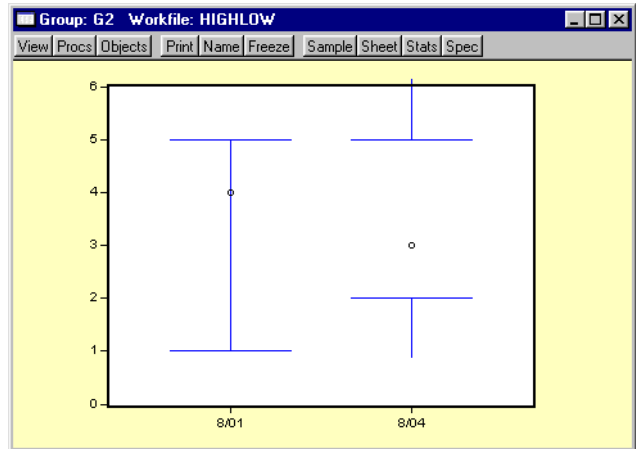


You can display only symbols in XY Lines without connecting the points. Click anywhere in the background of the XY Line view and change the Line Graphs option from **Lines only** to **Symbols only**. Alternatively, you can connect the points in scatter diagrams by marking the **Connect points** option in the Graph Options dialog box.

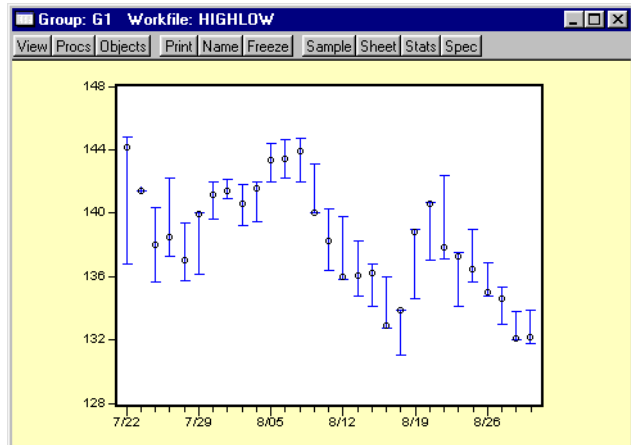
## High-Low(-Close)

This view plots the first three series in the group as a high-low-close chart. The first series is the “high” value and the second series is the “low” value. The high and low values are connected with a vertical line; if the high value is *below* the low value, the two have outside half-lines. The third series is the “close” value and is plotted as a small circle.

As the name suggests, this chart is commonly used by financial analysts to plot the daily high, low, and closing values of stock prices. The high-low-close chart for Microsoft is depicted below.

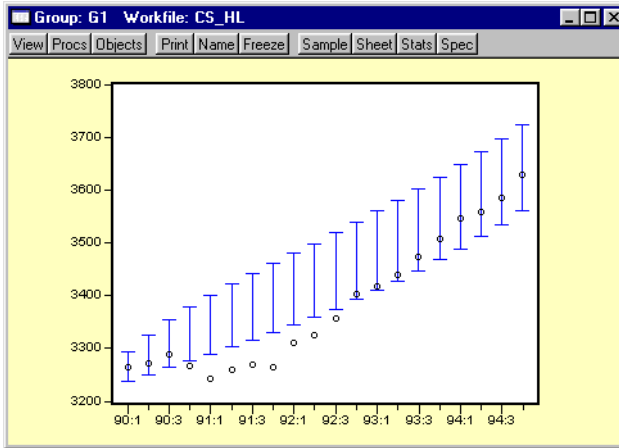


You can also use high-low charts to display forecast (or prediction) intervals. Suppose you saved the forecasts of consumption and its standard errors as series named CS\_F and CS\_SE (see Chapter 15 for a discussion of forecasting and computing fitted values). If you have actual data of CS for the forecast period, you can plot the approximate 95% prediction interval and the actual series in a high-low-close chart. First create a group (named say GR1) where the first series is the upper bound of the forecast interval, the second series is the lower bound, and the third series is the actual series.



```
group gr1 cs_f+2*cs_se cs_f-2*cs_se cs
```

Note that the order of the series in the group is important. Change the sample to the forecast period, double click on the group name GR1 and choose **View/Graph/High-Low**, which may look as follows:



This view clearly shows the eight quarters in which the forecasts failed to capture the drop in consumption.

### Pie Graph

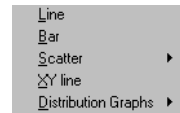
This view displays each observation as a pie chart, where the *percentage* of each series in the group is shown as a wedge in the pie.

If a series has a negative or missing value, the series will simply be dropped from the pie for that observation. You can label the observation number to each pie; double

click in the background of the pie chart and mark the **Label Pie** option in the Graph Options dialog.

### Multiple Graphs

While **Graph** views display all series in a single graph, **Multiple Graphs** views display a separate graph for each series in the group.



### Line and Bar Graphs

These views display a separate line graph or bar graph for each series in the group.

### Scatter

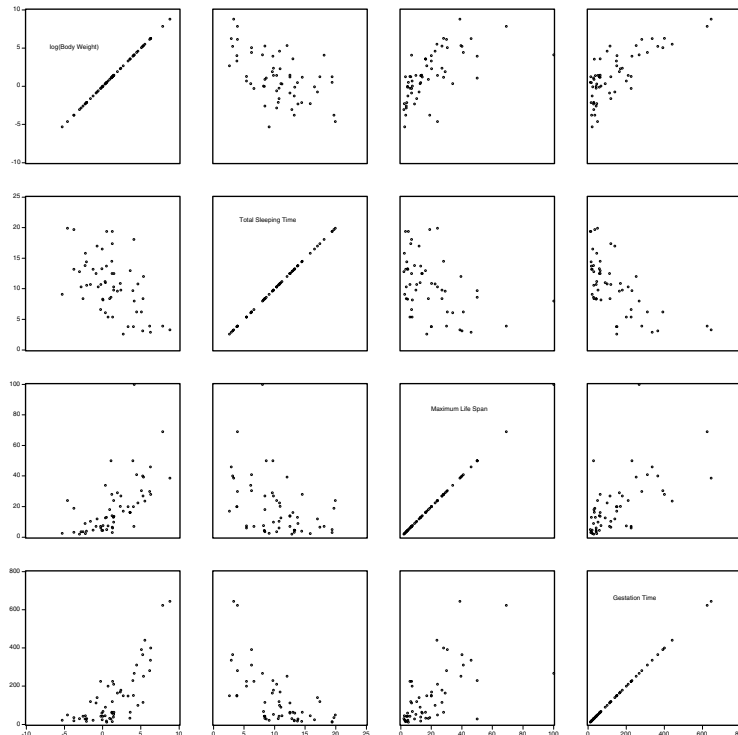
#### *First series against all*

This view displays scatter plots with the first series in the group on the horizontal axis and the remaining series on the vertical axis, each in a separate graph. If there are  $G$  series in the group,  $G^2$  scatter plots will be displayed.

#### *Scatter - Matrix of all pairs (SCATMAT)*

This view displays the *scatterplot matrix*, where scatter plots for all possible pairs of series in the group are displayed as a matrix. The important feature of the scatterplot matrix is that the scatter plots are arranged in such a way that plots in the same column share a common horizontal scale, while plots in the same row share a common vertical scale.

If there are  $G$  series in the group,  $G^2$  scatter plots will be displayed. The  $G$  plots on the main diagonal all lie on a 45 degree line, showing the distribution of the corresponding series on the 45 degree line. The  $G \cdot (G - 1)/2$  scatters below and above the main diagonal are the same; they are repeated so that we can scan the plots both across the rows and across the columns.



Here is a scatter plot matrix copy-and-pasted into a wordprocessor document. Note that the resolution of the scatter plot matrix deteriorates quickly as the number of series in the group increases. You may want to freeze the view and modify the graph by moving the axis labels into the scatters on the main diagonal. You can also save more space by moving each scatter close to each other. Simply click on the graph and drag each graph closer.

## XY line

This view plots the XY line graph of the first series on the horizontal X-axis and each of the remaining series on the vertical Y-axis in separate graphs. See the XY line view for **Graph** for more information on XY line graphs. If there are  $G$  series in the group,  $G \times 1$  XY line graphs will be displayed.

## Distribution Graphs

### CDF-Survivor-Quantile

This view displays the empirical cumulative distribution functions (CDF), survivor functions, and quantiles of each series in the group. These are identical to the series **CDF-Survivor-Quantile** view; see Chapter 9, in “CDF-Survivor-Quantile” on page 197 for a detailed description of how these graphs are computed and the available options.

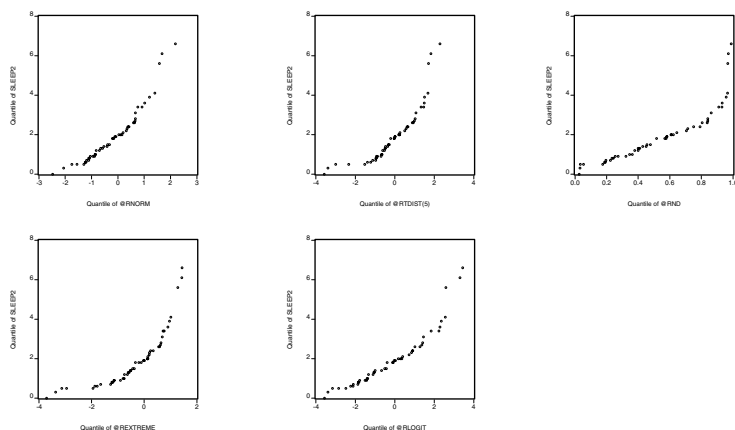
### Quantile-Quantile

This view plots the quantiles of each series against the quantiles of a specified distribution or the empirical quantiles of another series. QQ-plots are explained in detail in Chapter 9, page 199.

One useful application of group QQ-plots is to form a group of simulated series from different distributions and plot them against the quantiles of the series of interest. This way you can view, at a glance, the QQ-plots against various distributions. Suppose you want to know the distribution of the series SLEEP2. First, create a group containing random draws from the candidate distributions. For example,

```
group dist @rnorm @rtldist(5) @rextreme @rlogit @rnd
```

creates a group named DIST that contains simulated random draws from the standard normal,  $t$  with 5 degrees of freedom, extreme value, logistic, and uniform distributions. Open the group DIST, choose **View/Multiple Graphs/Distribution Graphs/Quantile-Quantile**, select the **Series or Group** option and type in the name of the series SLEEP2 in the field of the QQ Plot dialog box.



The quantiles of SLEEP2 are plotted on the vertical axis of each graph. (We moved one of the graphs to make the plots a bit easier to see.) The QQ-plot of the underlying distribution should

lie on a straight line. In this example, none of the QQ-plots lie on a line, indicating that the distribution of SLEEP2 does not match any of those in the group DIST.

## Descriptive Stats

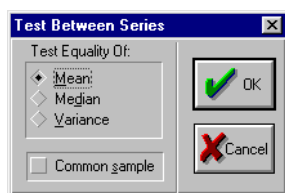
This view displays the summary statistics of each series in the group. Details for each statistic are provided in Chapter 7.

**Common Sample** computes the statistics using observations for which there are no missing values in any of the series in the group (casewise deletion of observations).

**Individual Samples** computes the statistics using all nonmissing observations for each series (listwise deletion).

The two views are identical if there are no missing values, or if every series has missing observations at the same observation numbers.

## Tests of Equality



This view tests the null hypothesis that all series in the group have the same mean, median, or variance. All of these tests are described in detail in Chapter 3, “Simple Hypothesis Tests” on page 148.

The **Common sample** option uses only observations for which none of the series in the group has missing values.

## Example

As an illustration, we demonstrate the use of this view to test for groupwise heteroskedasticity. Suppose we use data for seven countries over the period 1950–1992 and estimate a pooled OLS model (see Chapter 22). To test whether the residuals from this pooled regression are groupwise heteroskedastic, we test the equality of the variances of the residuals for each country.

First, save the residuals from the pooled OLS regression and make a group of the residuals corresponding to each country. This is most easily done by estimating the pooled OLS regression using a pool object and saving the residuals by selecting **Procs/Make Residuals** in the pool object menu or toolbar.

Next, open a group containing the residual series. One method is to highlight each residual series with the right mouse button, double click in the highlighted area and select **Open Group**. Alternatively, you can type show, followed by the names of the residual series, in the command window.

Select **View/Tests of Equality...**, and choose the **Variance** option in the Test Between Series dialog box.

## Test for Equality of Variances between Series

Date: 10/20/97 Time: 15:24

Sample: 1950 1992

Included observations: 43

Method	df	Value	Probability
Bartlett	6	47.65089	1.39E-08
Levene	(6, 287)	5.947002	7.15E-06
Brown-Forsythe	(6, 287)	4.603232	0.000176

## Category Statistics

Variable	Count	Std. Dev.	Mean Abs. Mean Diff.	Mean Abs. Median Diff.
RESID_CAN	42	387.3328	288.2434	275.5092
RESID_FRA	42	182.4492	143.0463	140.4258
RESID_GER	42	224.5817	169.6377	167.0994
RESID_ITA	42	173.4625	132.1824	131.2676
RESID_JAP	42	230.4443	185.5166	185.5166
RESID_UK	42	218.8625	159.4564	157.8945
RESID_US	42	340.9424	271.5252	265.4067
All	294	263.4411	192.8011	189.0171

Bartlett weighted standard deviation: 262.1580

The test statistics decisively reject the null hypothesis of equal variance of the residuals across countries, providing strong evidence of the presence of groupwise heteroskedasticity. You may want to adjust the denominator degrees of freedom to take account of the number of estimated parameters in the regression. The tests are, however, consistent even without the degrees of freedom adjustment.

See Chapter 7, page 144, for a description of the test statistics for this hypothesis test, and page 148 for the tests for equality of means and medians.

## N-Way Tabulation

This view classifies the observations in the current sample into cells defined by the series in the group. You can display the cell counts in various forms and examine statistics for independence among the series in the group. **Select View/N-Way Tabulation...** which opens the tabulation dialog.

Many of the settings will be familiar from our discussion of one-way tabulation in Chapter 7.

## Group into Bins If

If one or more of the series in the group is continuous and takes many distinct values, the number of cells becomes excessively large. This option provides you two ways to automatically bin the values of the series into subgroups.

**Number of values** option bins the series if the series takes more than the specified number of distinct values.

**Average count** option bins the series if the average count for each distinct value of the series is less than the specified number.

**Maximum number of bins** specifies the approximate maximum number of subgroups to bin the series. The number of bins may be chosen to be smaller than this number in order to make the bins approximately the same size.

The default setting is to bin a series into approximately 5 subgroups if the series takes more than 100 distinct values or if the average count is less than 2. If you do not want to bin the series, unmark both options.

## NA Handling

By default, EViews drops observations from the contingency table where any of the series in the group has a missing value. **Treat NA as category** option includes all observations and counts NAs in the contingency table as an explicit category.

## Layout

This option controls the display style of the tabulation. The **Table** mode displays the categories of the first two series in  $r$  by  $c$  tables for each category of the remaining series in the group.

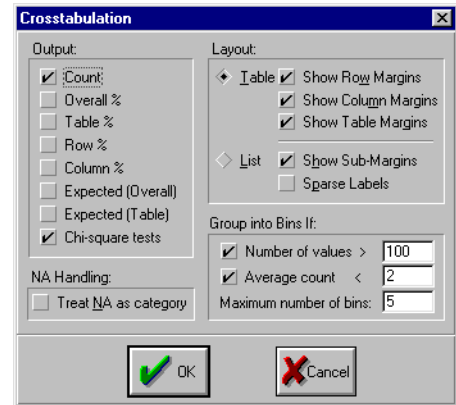
The **List** mode displays the table in a more compact, hierarchical form. The **Sparse Labels** option omits repeated category labels to make the list less cluttered. Note that some of the conditional  $\chi^2$  statistics are not displayed in list mode.

## Output

To understand the options for output, consider a group with three series. Let  $i, j, k$ , index the bin of the first, second, and third series, respectively. The number of observations in the

$(i, j, k)$ -th cell is denoted as  $n_{ijk}$  with a total of  $N = \sum_i \sum_j \sum_k n_{ijk}$  observations.

**Overall %** is the percentage of the total number of observations accounted for by the cell



count.

**Table %** is the percentage of the total number of observations in the conditional table accounted for by the cell count.

**Row %** is the percentage of the number of observations in the row accounted for by the cell count.

**Column %** is the percentage of the number of observations in the column accounted for by the cell count.

The *overall* expected count in the  $(i, j, k)$ -th cell is the number expected if *all* series in the group were independent of each other. This expectation is estimated by

$$\hat{n}_{ijk} = \left( \sum_i n_{ijk} / N \right) \left( \sum_j n_{ijk} / N \right) \left( \sum_k n_{ijk} / N \right) \cdot N . \quad (8.2)$$

The *table* expected count  $\tilde{n}_{ijk}$  is estimated by computing the expected count for the conditional table. For a given table, this expected value is estimated by:

$$\tilde{n}_{ijk^*} = \left( \sum_i n_{ijk^*} / N_{k^*} \right) \left( \sum_j n_{ijk^*} / N_{k^*} \right) \cdot N_{k^*} \quad (8.3)$$

where  $N_{k^*}$  is the total number of observations in the  $k^*$  table.

### Chi-square tests

If you select the **Chi-square tests** option, EViews reports  $\chi^2$  statistics for testing the independence of the series in the group. The test statistics are based on the distance between the actual cell count and the count expected under independence.

#### *Overall (unconditional) independence among all series in the group*

EViews reports the following two test statistics for overall independence among all series in the group:

$$\text{Pearson } \chi^2 = \sum_{i, j, k} \frac{(\hat{n}_{i, j, k} - n_{i, j, k})^2}{\hat{n}_{i, j, k}} \quad (8.4)$$

$$\text{Likelihood ratio} = 2 \sum_{i, j, k} n_{i, j, k} \log \left( \frac{n_{i, j, k}}{\hat{n}_{i, j, k}} \right)$$

where  $n_{ijk}$  and  $\hat{n}_{ijk}$  are the actual and overall expected count in each cell. Under the null hypothesis of independence, the two statistics are asymptotically distributed  $\chi^2$  with  $(I-1)(J-1)(K-1)$  degrees of freedom where  $I, J, K$  are the number of categories for each series.

These test statistics are reported at the top of the contingency table:



Tabulation of LWAGE and UNION and MARRIED

Date: 10/20/97 Time: 15:27

Sample: 1 1000

Included observations: 1000

---

Tabulation Summary

---

<u>Variable</u>	<u>Categories</u>
LWAGE	5
UNION	2
MARRIED	2
Product of Categories	20

<u>Test Statistics</u>	<u>df</u>	<u>Value</u>	<u>Prob</u>
Pearson X2	4	174.5895	0.000000
Likelihood Ratio G2	4	167.4912	0.000000

---

WARNING: Expected value is less than 5 in 40.00% of cells (8 of 20).

---

In this group there are three series LWAGE, UNION, and MARRIED, each with  $I = 5$ ,  $J = 2$ , and  $K = 2$  categories. Note the WARNING message: if there are many cells with expected value less than 5, the small sample distribution of the test statistic under the null hypothesis may deviate considerably from the asymptotic  $\chi^2$  distribution.

*Conditional independence between series in the group*

If you display in table mode, EViews presents measures of association for *each* conditional table. These measures are analogous to the correlation coefficient; the larger the measure, the larger the association between the row series and the column series in the table. In addition to the Pearson  $\chi^2$  for the table, the following three measures of association are reported:

$$\text{Phi coefficient} = \sqrt{\tilde{\chi}^2 / \tilde{N}} \quad (8.5)$$

$$\text{Cramers V} = \sqrt{\tilde{\chi}^2 / ((\min\{r, c\} - 1)\tilde{N})} \quad (8.6)$$

$$\text{Contingency coefficient} = \sqrt{\tilde{\chi}^2 / (\tilde{\chi}^2 + N)} \quad (8.7)$$

where  $\min(r, c)$  is the smaller of the number of row categories  $r$  or column categories  $c$  of the table, and  $\tilde{N}$  is the number of observations in the table. Note that all three measures are bounded between 0 and 1, a higher number indicating a stronger relation between the two series in the table. While the correlation coefficient only measures the linear association between two series, these nonparametric measures are robust to departures from linearity.

Table 1: Conditional table for MARRIED=0:

Count		UNION		Total
		0	1	
LWAGE	[0, 1)	0	0	0
	[1, 2)	167	8	175
	[2, 3)	121	44	165
	[3, 4)	17	2	19
	[4, 5)	0	0	0
Total		305	54	359

Measures of Association	Value
Phi Coefficient	0.302101
Cramer's V	0.302101
Contingency Coefficient	0.289193

Table Statistics	df	Value	Prob
Pearson X2	2	32.76419	7.68E-08
Likelihood Ratio G2	2	34.87208	2.68E-08

Note: Expected value is less than 5 in 16.67% of cells (1 of 6).

Bear in mind that these measures of association are computed for each two-way table. The conditional tables are presented at the top, and the unconditional tables are reported at the bottom of the view.

## Correlations, Covariances, and Correlogram

**Correlations** and **Covariances** display the correlation and covariance matrix of the series in the group. Observations for which any one of the series has missing data are excluded from the calculation of correlations and covariances.

**Correlogram** displays the autocorrelations and partial autocorrelations of the first series in the group. See Chapter 7, page 156, for a description of the correlogram view.

## Cross Correlations

This view displays the cross correlations of the first two series in the group. The cross correlations between the two series  $x$  and  $y$  are given by

$$r_{xy}(l) = \frac{c_{xy}(l)}{\sqrt{c_{xx}(0)} \cdot \sqrt{c_{yy}(0)}}, \quad \text{where } l = 0, \pm 1, \pm 2, \dots \quad (8.8)$$

and

$$c_{xy}(l) = \begin{cases} \sum_{t=1}^{T-l} ((x_t - \bar{x})(y_{t+l} - \bar{y}))/T & l = 0, 1, 2, \dots \\ \sum_{t=1}^{T+l} ((y_t - \bar{y})(x_{t-l} - \bar{x}))/T & l = 0, -1, -2, \dots \end{cases} \quad (8.9)$$

Note that, unlike the autocorrelations, cross correlations are not necessarily symmetric around lag 0.

The dotted lines in the cross correlograms are the approximate two standard error bounds computed as  $\pm 2/\sqrt{T}$ .

## Cointegration Test

This view carries out the Johansen test for whether the series in the group are cointegrated or not. Chapter 20 discusses the Johansen test in detail and how to interpret the test results.

## Granger Causality

Correlation does not necessarily imply causation in any meaningful sense of that word. The econometric graveyard is full of magnificent correlations, which are simply spurious or meaningless. Interesting examples include a positive correlation between teachers' salaries and the consumption of alcohol and a superb positive correlation between the death rate in the UK and the proportion of marriages solemnized in the Church of England. Economists debate correlations which are less obviously meaningless.

The Granger (1969) approach to the question of whether  $x$  causes  $y$  is to see how much of the current  $y$  can be explained by past values of  $y$  and then to see whether adding lagged values of  $x$  can improve the explanation.  $y$  is said to be Granger-caused by  $x$  if  $x$  helps in the prediction of  $y$ , or equivalently if the coefficients on the lagged  $x$ 's are statistically significant. Note that two-way causation is frequently the case;  $x$  Granger causes  $y$  and  $y$  Granger causes  $x$ .

It is important to note that the statement " $x$  Granger causes  $y$ " does not imply that  $y$  is the effect or the result of  $x$ . Granger causality measures precedence and information content but does not by itself indicate causality in the more common use of the term.

When you select the **Granger Causality** view, you will first see a dialog box asking for the number of lags to use in the test regressions. In general it is better to use more rather than fewer lags, since the theory is couched in terms of the relevance of all past information. You should pick a lag length,  $l$ , that corresponds to reasonable beliefs about the longest time over which one of the variables could help predict the other.

EViews runs bivariate regressions of the form

$$\begin{aligned}y_t &= \alpha_0 + \alpha_1 y_{t-1} + \dots + \alpha_l y_{t-l} + \beta_1 x_{t-1} + \dots + \beta_l x_{t-l} + \epsilon_t \\x_t &= \alpha_0 + \alpha_1 x_{t-1} + \dots + \alpha_l x_{t-l} + \beta_1 y_{t-1} + \dots + \beta_l y_{t-l} + u_t\end{aligned}\quad (8.10)$$

for all possible pairs of  $(x, y)$  series in the group. The reported  $F$ -statistics are the Wald statistics for the joint hypothesis:

$$\beta_1 = \beta_2 = \dots = \beta_l = 0 \quad (8.11)$$

for each equation. The null hypothesis is therefore that  $x$  does *not* Granger-cause  $y$  in the first regression and that  $y$  does *not* Granger-cause  $x$  in the second regression. The test results are given by

Pairwise Granger Causality Tests

Date: 10/20/97 Time: 15:31

Sample: 1946:1 1995:4

Lags: 4

Null Hypothesis:	Obs	F-Statistic	Probability
GDP does not Granger Cause CS	189	1.39156	0.23866
CS does not Granger Cause GDP		7.11192	2.4E-05

For this example, we cannot reject the hypothesis that GDP does not Granger cause CS but we do reject the hypothesis that CS does not Granger cause GDP. Therefore it appears that Granger causality runs one-way from CS to GDP and not the other way.

If you want to run Granger causality tests with other exogenous variables (e.g. seasonal dummy variables or linear trends) or if you want to carry out likelihood ratio (LR) tests, run the test regressions directly using equation objects.

## Label

This view displays descriptions of the group. You can edit any of the field cells in the label view, except the Last Update cell which shows the date/time the group was last modified.

Name is the group name as it appears in the workfile; you can rename your group by editing this cell. If you fill in the Display Name cell, EViews will use this name in some of the tables and graphs of the group view. Unlike Names, Display Names may contain spaces and preserve capitalization (upper and lower case letters).

See Chapter 6 for a discussion of the label fields and their use in database searches.

## Group Procedures

There are two procedures available for groups.

**Make Equation...** opens an Equation Specification dialog box with the first series in the

group listed as the dependent variable and the remaining series as the regressors, including a constant term *C*. You can modify the specification as you wish.

**Make Vector Autoregression...** opens an Unrestricted Vector Autoregression dialog box, where all series in the group are listed as endogenous variables in the VAR. You can modify the VAR specification as you wish; see Chapter 20 for a discussion of specifying and estimating VARs in EViews.

## Commands

You can type the following commands as an alternative to choosing from menus in the group window toolbar. The general rule is to follow the name of the group with a dot and command name for the view or procedure with options specified in parentheses. For example, to view the simple scatter diagram of a group named GRP1, type

```
grp1.scat
```

To test whether the medians of all the series in group GP\_WAGE are equal, type

```
gp_wage.testbet(med)
```

To plot the cross correlogram of the first two series in group GRP\_MACRO up to 12 lags, type

```
grp_macro.cross(12)
```



## Chapter 9. Statistical Graphs Using Series and Groups

EViews provides several methods for exploratory data analysis. In Chapter 7 we document several graph views that may be used to characterize the distribution of a series. This chapter describes several bivariate scatterplot views which allow you to fit lines using parametric, and nonparametric procedures.

These views, which involve relatively complicated calculations or have a number of specialized options, are documented in detail below. While the discussion sometimes involves some fairly technical issues, you should not feel as though you need to master all of the details to use these views. The graphs correspond to familiar concepts, and are designed to be simple and easy to understand visual displays of your data. The EViews default settings should be sufficient for all but the most specialized of analyses. Feel free to explore each of the views, clicking on **OK** to accept the default settings.

### Distribution Graphs of Series

The view menu of a series lists three graphs that characterize the empirical distribution of the series.

CDF - Survivor - Quantile...  
Quantile - Quantile...  
Kernel Density...

#### CDF-Survivor-Quantile

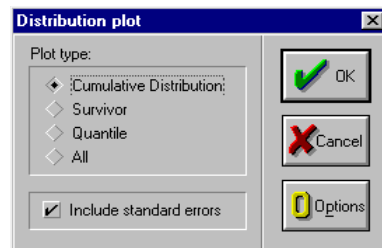
This view plots the empirical cumulative distribution, survivor, and quantile functions of the series together with the plus or minus two standard error bands. Select **View/Distribution Graphs/CDF-Survivor-Quantile...**

- The **Cumulative Distribution** option plots the empirical cumulative distribution function (CDF) of the series. The CDF is the probability of observing a value from the series not exceeding a specified value  $r$

$$F_x(r) = \Pr(x \leq r) . \quad (9.1)$$

- The **Survivor** option plots the empirical survivor function of the series. The survivor function gives the probability of observing a value from the series at least as large as some specified value  $r$  and is equal to one minus the CDF:

$$S_x(r) = \Pr(x > r) = 1 - F_x(r) . \quad (9.2)$$



- The **Quantile** option plots the empirical quantiles of the series. For  $0 < q < 1$ , the  $q$ -th quantile  $x_{(q)}$  of  $x$  is a number such that

$$\begin{aligned} \Pr(x \leq x_{(q)}) &\leq q \\ \Pr(x \geq x_{(q)}) &\leq 1 - q \end{aligned} \quad (9.3)$$

The quantile is the inverse function of the CDF; graphically, the quantile can be obtained by flipping the horizontal and vertical axis of the CDF.

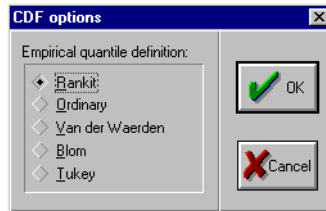
- The **All** option plots the CDF, survivor, and quantiles.

### Standard Errors

The **Include standard errors** option plots the approximate 95% confidence intervals together with the empirical distribution functions. The methodology for computing these intervals is described in detail in Conover (1980, pp. 114–116). Note that, using this approach, we do not compute confidence intervals for the quantiles corresponding to the first and last few order statistics.

### Options

EViews provides several methods of computing the empirical CDF used in the CDF and quantile computations.



Given a total of  $N$  observations, the CDF for value  $r$  is estimated as

Rankit (default)	$(r - 1/2)/N$
Ordinary	$r/N$
Van der Waerden	$r/(N + 1)$
Blom	$(r - 3/8)/(N + 1/4)$
Tukey	$(r - 1/3)/(N + 1/3)$

The various methods differ in how they adjust for the non-continuity of the CDF computation. The differences between these alternatives will become negligible as the sample size  $N$  grows.



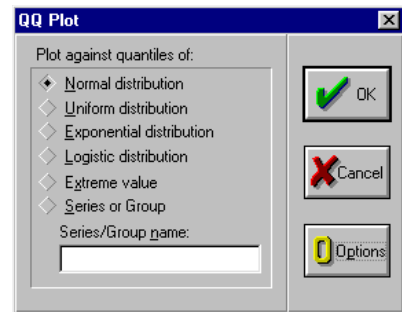
## Quantile-Quantile

The quantile-quantile (QQ)-plot is a simple yet powerful tool for comparing two distributions (Cleveland, 1994). This view plots the quantiles of the chosen series against the quantiles of another series or a theoretical distribution. If the two distributions are the same, the QQ-plot should lie on a straight line. If the QQ-plot does not lie on a straight line, the two distributions differ along some dimension. The pattern of deviation from linearity provides an indication of the nature of the mismatch.

When you select **View/Distribution Graphs/Quantile-Quantile...**, the QQ Plot dialog box appears.

You can plot against the quantiles of the following theoretical distributions:

- **Normal.** Bell-shaped and symmetric distribution.
- **Uniform.** Rectangular density function. Equal probabilities associated with any fixed interval size in the support.
- **Exponential.** The unit exponential is a positively skewed distribution with a long right tail.
- **Logistic.** This symmetric distribution is similar to the normal, except that it has longer tails than the normal.
- **Extreme value.** The Type-I (minimum) extreme value is a negatively skewed distribution with a long left tail—it is very close to a lognormal distribution.



You can also plot against the quantiles of any series in your workfile. Type the names of the series or groups in the edit box, and select **Series** or **Group**. EViews will compute a QQ-plot against each series in the list. You can use this option to plot against the quantiles of a simulated series from any distribution; see the example below.

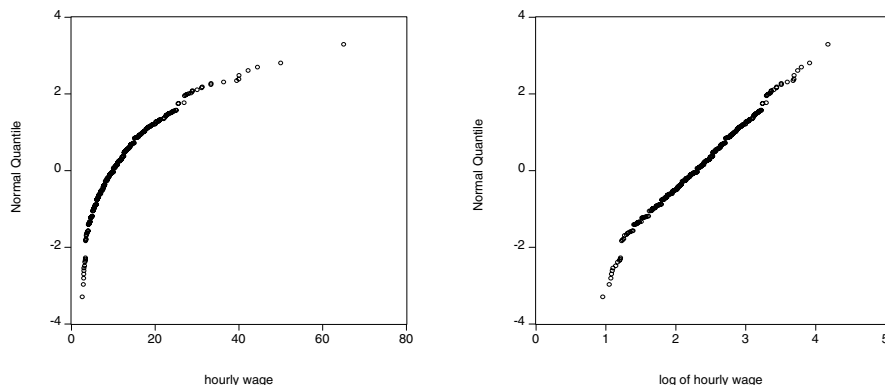
The **Options** button provides you with several methods for computing the empirical quantiles. The options are explained in the CDF-Survivor-Quantile section above; the choice should not make much difference unless the sample is very small.

For additional details, see Cleveland (1994), or Chambers, et al. (1983, Chapter 6).

## Illustration

Labor economists typically estimate wage earnings equations with the log of wage on the left-hand side instead of the wage itself. This is because the log of wage has a distribution more close to the normal than the wage, and classical small sample inference procedures are more likely to be valid. To check this claim, we can plot the quantiles of the wage and log of wage against

those from the normal distribution. Highlight the series, double click, select **View/Distribution Graphs/Quantile-Quantile...**, and choose the (default) **Normal distribution** option



If the distributions of the series on the vertical and horizontal axes match, the plots should lie on a straight line. The two plots clearly indicate that the log of wage has a distribution closer to the normal than the wage.

The concave shape of the QQ-plot for the wage indicates that the distribution of the wage series is positively skewed with a long right tail. If the shape were convex, it would indicate that the distribution is negatively skewed.

The QQ-plot for the log of wage falls nearly on a straight line except at the left end, where the plot curves downward, so that log wages are somewhat platykurtic. QQ-plots that fall on a straight line in the middle but curve upward at the left end and curve downward at the right end indicate that the distribution is leptokurtic and has a thicker tail than the normal distribution. If the plot curves downward at the left, and upward at the right, it is an indication that the distribution is platykurtic and has a thinner tail than the normal distribution.

If you want to compare your series with a distribution not in the option list, you can use the random number generator in EViews and plot against the quantiles of the simulated series from the distribution. For example, suppose we wanted to compare the distribution of the log of wage with the  $F$ -distribution with 10 numerator degrees of freedom and 50 denominator degrees of freedom. First generate a random draw from  $F(10,50)$  by the command

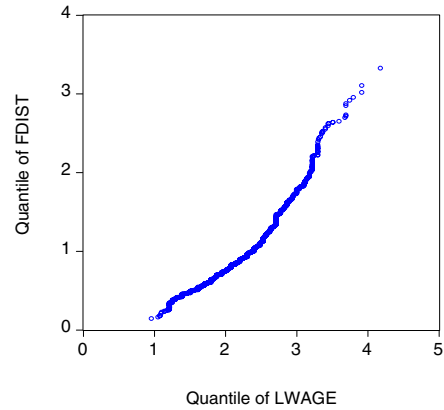
```
series fdist=@rfdist(10,50)
```

Then highlight the log of wage series, double click, select **View/Distribution Graphs/Quantile-Quantile...**, and choose the **Series or Group** option and type in the name of the simulated series (in this case `fdist`)

The plot is slightly convex, indicating that the distribution of the log of wage is slightly negatively skewed compared to the  $F(10,50)$ .

### Kernel Density

This view plots the kernel density estimate of the distribution of the series. The simplest nonparametric density estimate of a distribution of a series is the histogram. You can view the histogram by selecting **View/Descriptive Statistics/Histogram and Stats**. The histogram, however, is sensitive to the choice of origin and is not continuous.

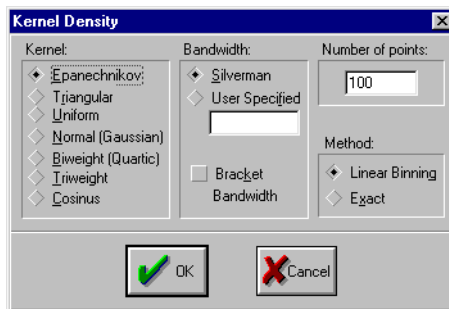


The kernel density estimator replaces the “boxes” in a histogram by “bumps” that are smooth (Silverman 1986). Smoothing is done by putting less weight on observations that are further from the point being evaluated. More technically, the kernel density estimate of a series  $X$  at a point  $x$  is estimated by:

$$f(x) = \frac{1}{Nh} \sum_{i=1}^N K\left(\frac{x - X_i}{h}\right), \quad (9.4)$$

where  $N$  is the number of observations,  $h$  is the bandwidth (or smoothing parameter) and  $K$  is a kernel function that integrates to one.

When you choose **View/Distribution Graphs/Kernel Density...**, the Kernel Density dialog appears.



To display the kernel density estimates, you need to specify the following:

- **Kernel.** The kernel function is a weighting function that determines the shape of the bumps. EViews provides the following options for the kernel function  $K$ :

Epanechnikov (default)	$\frac{3}{4}(1 - u^2)I( u  \leq 1)$
Triangular	$(1 -  u )(I( u  \leq 1))$
Uniform (Rectangular)	$\frac{1}{2}(I( u  \leq 1))$
Normal (Gaussian)	$\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}u^2\right)$
Biweight (Quartic)	$\frac{15}{16}(1 - u^2)^2 I( u  \leq 1)$
Triweight	$\frac{35}{32}(1 - u^2)^3 I( u  \leq 1)$
Cosinus	$\frac{\pi}{4}\cos\left(\frac{\pi}{2}u\right)I( u  \leq 1)$

where  $u$  is the argument of the kernel function and  $I$  is the indicator function that takes a value of one if its argument is true, and zero otherwise.

- **Bandwidth.** The bandwidth  $h$  controls the smoothness of the density estimate; the larger the bandwidth, the smoother the estimate. Bandwidth selection is of crucial importance in density estimation (Silverman 1986), and various methods have been suggested in the literature. The **Silverman** option (default) uses a data-based automatic bandwidth

$$h = 0.9kN^{-1/5}\min(s, R/1.34) \quad (9.5)$$

where  $N$  is the number of observations,  $s$  is the standard deviation, and  $R$  is the interquartile range of the series (Silverman 1986, equation 3.31). The factor  $k$  is a canonical bandwidth-transformation that differs across kernel functions (Marron and Nolan 1989; Härdle 1991). The canonical bandwidth-transformation adjusts the bandwidth so that the automatic density estimates have roughly the same amount of smoothness across various kernel functions.

To specify a bandwidth of your choice, mark **User Specified** option and type a nonnegative number for the bandwidth in the field box. Although there is no general rule for the appropriate choice of the bandwidth, Silverman (1986, section 3.4) makes a case for under-smoothing by choosing a somewhat small bandwidth, since it is easier for the eye to smooth than it is to unsmooth.

The **Bracket Bandwidth** option allows you to investigate the sensitivity of your estimates to variations in the bandwidth. If you choose to bracket the bandwidth, EViews plots three density estimates using bandwidths  $0.5h$ ,  $h$ , and  $1.5h$ .

- **Number of Points.** You must specify the number of points  $M$ , at which you will evaluate the density function. The default is  $M = 100$  points. Suppose the minimum and maximum value to be considered are given by  $X_L$  and  $X_U$ , respectively. Then  $f(x)$  is evaluated at  $M$  equi-spaced points given by:

$$x_i = X_L + i \cdot \left( \frac{X_U - X_L}{M} \right), \text{ for } i = 0, 1, \dots, M-1. \quad (9.6)$$

EViews selects the lower and upper evaluation points by extending the minimum and maximum values of the data by two (for the normal kernel) or one (for all other kernels) bandwidth units.

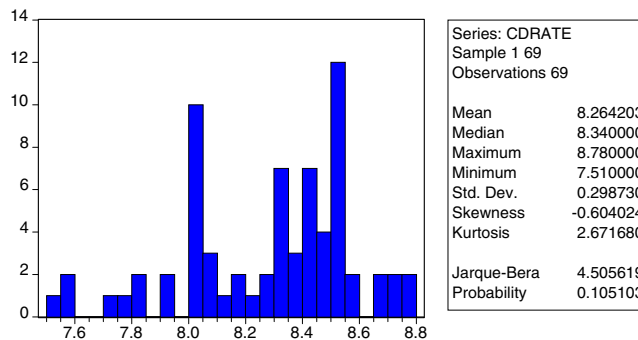
- **Method.** By default, EViews utilizes the **Linear Binning** approximation algorithm of Fan and Marron (1994) to limit the number of evaluations required in computing the density estimates. For large samples, the computational savings are substantial.

The **Exact** option evaluates the density function using all of the data points for each  $X_j$ ,  $j = 1, 2, \dots, N$  for each  $x_i$ . The number of kernel evaluations is therefore of order  $O(NM)$ , which, for large samples, may be quite time-consuming.

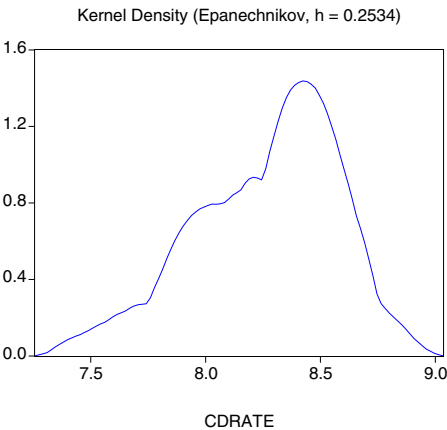
Unless there is a strong reason to compute the exact density estimate or unless your sample is very small, we recommend that you use the binning algorithm.

### Illustration

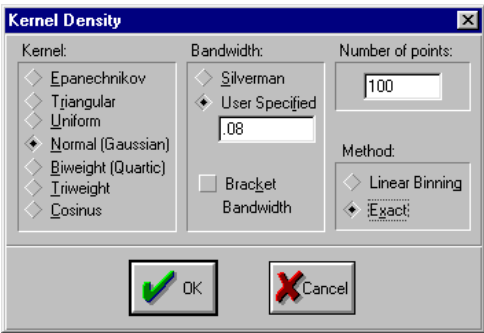
As an illustration of kernel density estimation, we use the three month CD rate data for 69 Long Island banks and thrifts used in Simonoff (1996). The histogram of the CD rate looks as follows:



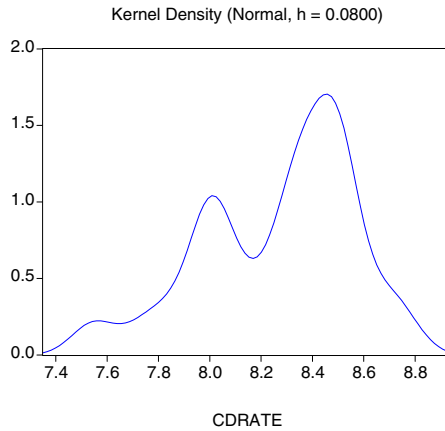
This histogram is a very crude estimate of the distribution of CD rates and does not provide us with much information about the underlying distribution. To view the kernel density estimate, select **View/Distribution Graphs/Kernel Density....** The default options produced the following view:



This density estimate seems to be oversmoothed. Simonoff (1996, chapter 3) uses a Gaussian kernel with bandwidth 0.08. To replicate his results, select **View/Distribution Graphs/Kernel Density...**, and fill in the dialog box as follows:



Note that we select the **Exact** method option since there are only 69 observations to evaluate the kernel. The kernel density result is depicted below:



This density estimate has about the right degree of smoothing. Interestingly enough, this density has a trimodal shape with modes at the “focal” numbers 7.5, 8.0, and 8.5.

## Scatter Diagrams with Fit Lines

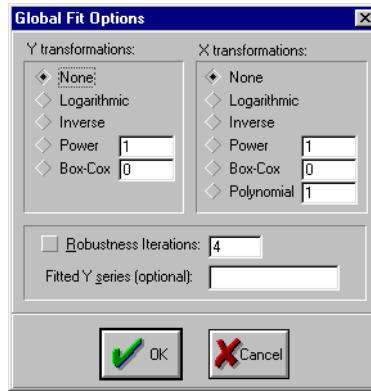
Simple Scatter  
Scatter with Regression  
Scatter with Nearest Neighbor Fit  
Scatter with Kernel Fit

The view menu of a group includes four variants of scatterplot diagrams. **Simple Scatter** plots a scatter diagram with the first series on the horizontal axis and the remaining series on the vertical axis.

The remaining three options, **Scatter with Regression**, **Scatter with Nearest Neighbor Fit**, and **Scatter with Kernel Fit** plot fitted lines for the second series.

### Scatter with Regression

This view fits a bivariate regression of transformations of the second series in the group Y on transformations of the first series in the group X (and a constant).



The following transformations of the series are available for the bivariate fit:

None	$y$	$x$
Logarithmic	$\log(y)$	$\log(x)$
Inverse	$1/y$	$1/x$
Power	$y^a$	$x^b$
Box-Cox	$(y^a - 1)/a$	$(x^b - 1)/b$
Polynomial	—	$1, x, x^2, \dots, x^b$

where you specify the parameters  $a$  and  $b$  in the edit field. Note that the Box-Cox transformation with parameter zero is the same as the log transformation.

- If any of the transformed values are not available, EViews returns an error message. For example, if you take logs of negative values, noninteger powers of nonpositive values, or inverses of zeros, EViews will stop processing and issue an error message.
- If you specify a high-order polynomial, EViews may be forced to drop some of the high order terms to avoid collinearity.

When you click **OK**, EViews displays a scatter diagram of the series together with a line connecting the fitted values from the regression. You may optionally save the fitted values as a series. Type a name for the fitted series in the **Fitted Y series** edit field.

### Robustness Iterations

The least squares method is very sensitive to the presence of even a few outlying observations. The **Robustness Iterations** option carries out a form of weighted least squares where outlying observations are given relatively less weight in estimating the coefficients of the regression.



For any given transformation of the series, the **Robustness Iteration** option carries out robust fitting with bisquare weights. Robust fitting estimates the parameters  $a$ ,  $b$  to minimize the weighted sum of squared residuals

$$\sum_{i=1}^N r_i (y_i - a - x_i b)^2 \quad (9.7)$$

where  $y_i$  and  $x_i$  are the transformed series and the bisquare robustness weights  $r$  are given by

$$r = \begin{cases} (1 - e_i^2 / (36m^2))^2 & \text{for } |e_i / 6m| < 1 \\ 0 & \text{otherwise} \end{cases} \quad (9.8)$$

where  $e_i = y_i - a - x_i b$  is the residual from the previous iteration (the first iteration weights are determined by the OLS residuals), and  $m$  is the median of  $|e_i|$ . Observations with large residuals (outliers) are given small weights when forming the weighted sum of squared residuals.

To choose robustness iterations, click on the check box for **Robustness Iterations** and specify an integer for the number of iterations.

See Cleveland (1993) for additional discussion.

## Scatter with Nearest Neighbor Fit

This view displays local polynomial regressions with bandwidth based on nearest neighbors. Briefly, for each data point in a sample, we fit a locally weighted polynomial regression. It is a local regression since we use only the subset of observations which lie in a neighborhood of the point to fit the regression model; it may be weighted so that observations further from the given data point are given less weight.

This class of regressions includes the popular *Loess* (also known as *Lowess*) techniques described by Cleveland (1993, 1994). Additional discussion of these techniques may be found in Fan and Gijbels (1996), and in Chambers, Cleveland, Kleiner, Tukey (1983).

### Method

You should choose between computing the local regression at each data point in the sample, or using a subsample of data points.

**Nearest Neighbor Fit Options**

**Specification:**

Bandwidth span: (fraction of data)

Polynomial degree:

☐ Bracket bandwidth span

**Options:**

☒ Local Weighting (Tricube)

☐ Robustness Iterations

# Iterations:

☐ Symmetric neighbors

**Method:**

☒ Exact (full sample)

☐ Cleveland subsampling

Approx points:

**Fitted series (optional):**

- **Exact (full sample)** fits a local regression at every data point in the sample.
- **Cleveland subsampling** performs the local regression at only a subset of points. You should provide the size of the subsample  $M$  in the edit box.

The number of points at which the local regressions are computed is approximately equal to  $M$ . The actual number of points will depend on the distribution of the explanatory variable.

Since the exact method computes a regression at every data point in the sample, it may be quite time consuming when applied to large samples. For samples with over 100 observations, you may wish to consider subsampling.

The idea behind subsampling is that the local regression computed at two adjacent points should differ by only a small amount. Cleveland subsampling provides an adaptive algorithm for skipping nearby points in such a way that the subsample includes all of the representative values of the regressor.

It is worth emphasizing that at each point in the subsample, EViews uses the entire sample in determining the neighborhood of points. Thus, each regression in the Cleveland subsample corresponds to an equivalent regression in the exact computation. For large data sets, the computational savings are substantial, with very little loss of information.

### Specification

For each point in the sample selected by the **Method** option, we compute the fitted value by running a local regression using data around that point. The **Specification** option determines the rules employed in identifying the observations to be included in each local regression, and the functional form used for the regression.

**Bandwidth span** determines which observations should be included in the local regressions. You should specify a number  $\alpha$  between 0 and 1. The span controls the smoothness of the local fit; a larger fraction  $\alpha$  gives a smoother fit. The fraction  $\alpha$  instructs EViews to include the  $[\alpha N]$  observations nearest to the given point, where  $[\alpha N]$  is  $100\alpha$  % of the total sample size, truncated to an integer.

Note that this standard definition of nearest neighbors implies that the number of points need not be symmetric around the point being evaluated. If desired, you can force symmetry by selecting the **Symmetric neighbors** option.

**Polynomial degree** specifies the degree of polynomial to fit in each local regression.

If you mark the **Bracket bandwidth span** option, EViews displays three nearest neighbor fits with spans of  $0.5\alpha$ ,  $\alpha$ , and  $1.5\alpha$ .

### Other Options

**Local Weighting (Tricube)** weights the observations of each local regression. The weighted regression minimizes the weighted sum of squared residuals

$$\sum_{i=1}^N w_i (y_i - a - x_i b_1 - x_i^2 b_2 - \dots - x_i^k b_k). \quad (9.9)$$

The tricube weights  $w$  are given by

$$w_i = \begin{cases} \left(1 - \left| \frac{d_i}{d(\lfloor \alpha N \rfloor)} \right| \right)^3 & \text{for } \left| \frac{d_i}{d(\lfloor \alpha N \rfloor)} \right| < 1 \\ 0 & \text{otherwise} \end{cases} \quad (9.10)$$

where  $d_i = |x_i - x|$  and  $d(\lfloor \alpha N \rfloor)$  is the  $\lfloor \alpha N \rfloor$ -th smallest such distance. Observations that are relatively far from the point being evaluated get small weights in the sum of squared residuals. If you turn this option off, each local regression will be unweighted with  $w_i = 1$  for all  $i$ .

**Robustness Iterations** iterates the local regressions by adjusting the weights to downweight outlier observations. The initial fit is obtained using weights  $w_i$ , where  $w_i$  is tricube if you choose **Local Weighting** and 1 otherwise. The residuals  $e_i$  from the initial fit are used to compute the robustness bisquare weights  $r_i$  as given on page 207. In the second iteration, the local fit is obtained using weights  $w_i r_i$ . We repeat this process for the user specified number of iterations, where at each iteration the robustness weights  $r_i$  are recomputed using the residuals from the last iteration.

**Symmetric Neighbors** forces the local regression to include the same number of observations to the left and to the right of the point being evaluated. This approach violates the definition, though not the spirit, of nearest neighbor regression.

To save the fitted values as a series; type a name in the **Fitted series** field box. If you have specified subsampling, EViews will linearly interpolate to find the fitted value of  $y$  for the actual value of  $x$ . If you have marked the **Bracket bandwidth span** option, EViews saves three series with  $\_L$ ,  $\_M$ ,  $\_H$  appended to the name, each corresponding to bandwidths  $0.5\alpha$ ,  $\alpha$ , and  $1.5\alpha$ , respectively.

Note that Loess is a special case of nearest neighbor fit, with a polynomial of degree 1, and local tricube weighting. The default EViews options are set to provide Loess estimation.

## Scatter with Kernel Fit

This view displays fits of local polynomial kernel regressions of the second series in the group Y on the first series in the group X. Both the nearest neighbor fit, described above, and the kernel fit are nonparametric regressions that fit local polynomials. The two differ in how they define “local” in the choice of bandwidth. The effective bandwidth in nearest neighbor regression varies, adapting to the observed distribution of the regressor. For the kernel fit, the bandwidth is fixed but the local observations are weighted according to a kernel function.

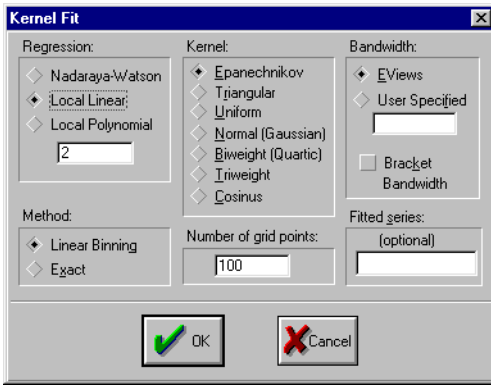
Extensive discussion may be found in Simonoff (1996), Hardle (1991), Fan and Gijbels (1996).

Local polynomial kernel regressions fit  $Y_i$  at each value  $x$ , by choosing the parameters  $\beta$  to minimize the weighted sum-of-squared residuals:

$$m(x) = \sum_{i=1}^N (Y_i - \beta_0 - \beta_1(x - X_i) + \dots - \beta_k(x - X_i)^k)^2 K\left(\frac{x - X_i}{h}\right) \quad (9.11)$$

where  $N$  is the number of observations,  $h$  is the bandwidth (or smoothing parameter), and  $K$  is a kernel function that integrates to one. Note that the minimizing estimates of  $\beta$  will differ for each  $x$ .

When you select the **Scatter with Kernel Fit** view, the Kernel Fit dialog appears.



You will need to specify the form of the local regression, the kernel, the bandwidth, and other options to control the fit procedure.

### Regression

Specify the order of polynomial  $k$  to be fit at each data point. The **Nadaraya-Watson** option sets  $k = 0$  and locally fits a constant at each  $x$ . **Local Linear** sets  $k = 1$  at each  $x$ . For higher order polynomials, mark the **Local Polynomial** option and type in an integer in the field box to specify the order of the polynomial.

### Kernel

The kernel is the function used to weight the observations in each local regression. EViews provides the option of selecting one of the following kernel functions:

Epanechnikov (default)	$\frac{3}{4}(1 - u^2)I( u  \leq 1)$
Triangular	$(1 -  u )(I( u  \leq 1))$
Uniform (Rectangular)	$\frac{1}{2}(I( u  \leq 1))$
Normal (Gaussian)	$\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}u^2\right)$
Biweight (Quartic)	$\frac{15}{16}(1 - u^2)^2 I( u  \leq 1)$

Triweight	$\frac{35}{32}(1-u^2)^3 I( u  \leq 1)$
Cosinus	$\frac{\pi}{4} \cos\left(\frac{\pi}{2}u\right) I( u  \leq 1)$

where  $u$  is the argument of the kernel function and  $I$  is the indicator function that takes a value of one if its argument is true, and zero otherwise

### Bandwidth

The bandwidth  $h$  determines the weights to be applied to observations in each local regression. The larger the  $h$ , the smoother the fit. By default, **EViews** arbitrarily sets the bandwidth to:

$$h = 0.15(X_U - X_L) \quad (9.12)$$

where  $(X_U - X_L)$  is the range of  $X$ .

For nearest neighbor bandwidths, see **Scatter with Nearest Neighbor Fit**.

To specify your own bandwidth, mark **User Specified** and enter a nonnegative number for the bandwidth in the edit box.

**Bracket Bandwidth** option fits three kernel regressions using bandwidths  $0.5h$ ,  $h$ , and  $1.5h$ .

### Number of grid points

You must specify for how many points  $M$  at which to evaluate the local polynomial regression. The default is  $M = 100$  points; you can specify any integer in the field. Suppose the range of the series  $X$  is  $[X_L, X_U]$ . Then the polynomial is evaluated at  $M$  equi-spaced points

$$x_i = X_L + i \cdot \left( \frac{X_U - X_L}{M} \right) \quad \text{for } i = 0, 1, \dots, M-1 \quad (9.13)$$

### Method

Given a number of evaluation points, EViews provides you with two additional computational options: exact computation and linear binning.

The **Exact** method performs a regression at each  $x_i$ , using all of the data points  $(X_j, Y_j)$ , for  $j = 1, 2, \dots, N$ . Since the exact method computes a regression at every grid point, it may be quite time consuming when applied to large samples. In these settings, you may wish to consider the linear binning method.

The **Linear Binning** method (Fan and Marron 1994) approximates the kernel regression by binning the raw data  $X_j$  fractionally to the two nearest evaluation points, prior to evaluating the

kernel estimate. For large data sets, the computational savings may be substantial, with virtually no loss of precision.

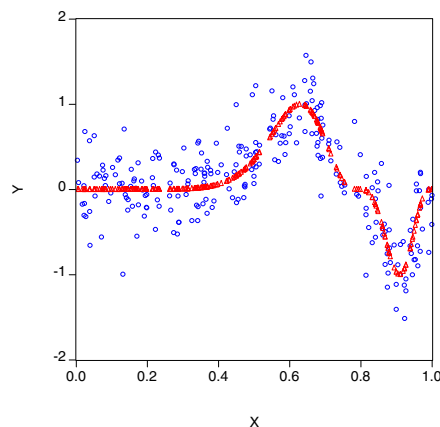
To save the fitted values as a series, type a name in the **Fitted Series** field box. EViews will save the fitted  $\hat{Y}$  to the series, linearly interpolating points computed on the grid, to find the appropriate value. If you have marked the **Bracket Bandwidth** option, EViews saves three series with  $\_L$ ,  $\_M$ ,  $\_H$  appended to the name, each corresponding to bandwidths  $0.5\alpha$ ,  $\alpha$ , and  $1.5\alpha$ , respectively.

### Example

As an example, we estimate a bivariate relation for a simulated data set used in Hardle (1991). The data were generated by

```
series x=rnd
series y=sin(2*pi*x^3)^3+nrnd*(0.1^.5)
```

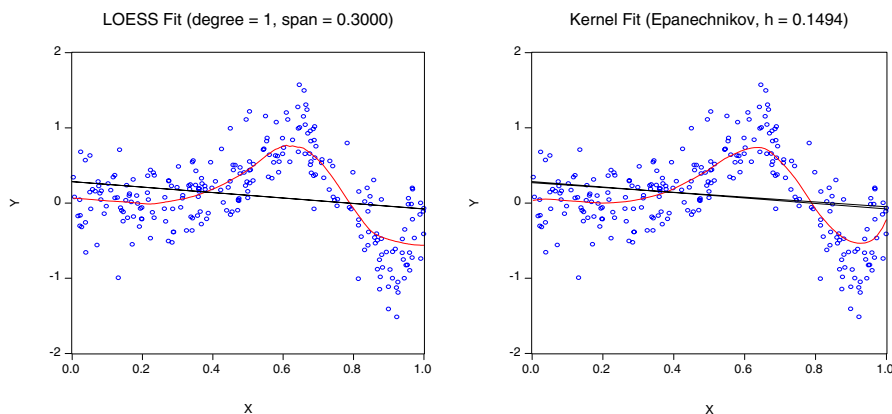
The simple scatter of  $Y$  and the "true" conditional mean of  $Y$  against  $X$  looks as follows:



The triangular shapes in the middle of the scatterplot trace out the "true" conditional mean of  $Y$ . Note that the true mean reaches a peak around  $x = 0.6$ , a valley around  $x = 0.9$ , and a saddle around  $x = 0.8$ .

To fit a nonparametric regression of  $Y$  on  $X$ , you first create a group containing the series  $Y$  and  $X$ . The order that you enter the series is important; the explanatory series variable must be the first series in the group. Highlight the series name  $X$  and then  $Y$ , double click in the highlighted area, select **Open Group**, and select **View/Graph/Scatter/Scatter with Nearest Neighbor Fit**, and repeat the procedure for **Scatter with Kernel Fit**.

The two fits, computed using the EViews default settings, are shown together with the linear regression line:



Both local regression lines seem to capture the peak, but the kernel fit is more sensitive to the upturn in the neighborhood of  $X = 1$ . Of course, the fitted lines change as we modify the options, particularly when we adjust the bandwidth  $h$  and window width  $\alpha$ .

## Commands

The command

```
lwage.cdfplot(a)
```

plots the CDF, quantile and survive functions for LWAGE.

```
lwage.kdensity(k=n)
```

plots the kernel density estimate of LWAGE using a normal kernel and the automatic bandwidth selection.

```
lwage.kdensity(k=e,b=.25)
```

plots the kernel density estimate using the Epanechnikov kernel, a bandwidth of .25, and bracketing.

```
group aa age lwage
aa.linefit(y1,x1)
```

creates a group AA containing the series LWAGE and AGE, and fits a regression line of log LWAGE on log AGE.

```
aa.linefit(y1,d=3)
```

fits log LWAGE to a third degree polynomial in AGE.

```
aa.nnfit
```

```
aa.kerfit
```

uses the default settings to generate scatterplots with (Loess) nearest-neighbor and kernel regression fits to the data in AA.



## Chapter 10. Graphs, Tables, and Text Objects

---

EViews objects (series, groups, equations, and so on) display their views as graphs, tables, and text. Views are dynamic. When the underlying object changes, or the active sample changes, so do the views. Often one would like to preserve the current view so that it does not change when the object changes. In EViews this is referred to as *freezing* the view. Freezing a graphical view creates a graph object, freezing a tabular view creates a table object, and freezing a text view creates a text object.

Frozen views form the basis of most presentation output and EViews provides extensive tools for customizing the appearance of these objects. This chapter describes the options available for controlling the appearance of graph, table, and text objects.

### Graphs

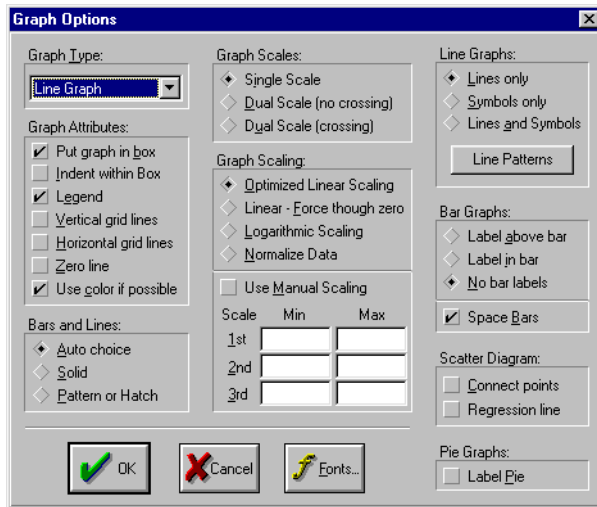
The appearance of both graphical views and graph objects can be customized extensively. However, while changes to a graph object are always stored permanently with the object, changes to a graphical view will often be lost when the view is redrawn (including when the object is closed and reopened, when the workfile sample is modified, or when the data underlying the object are changed). If you would like to customize a view for presentation purposes, you should first freeze the view as a graph object to ensure that your changes are not lost.

An EViews graph is made up of a number of elements: the plot area, the axes, the graph legend, and one or more pieces of added text or shading. To select one of these elements for editing, simply click in the area associated with it. A blue box will appear around the selected element. Once you have made your selection, you can click and drag to move the element around the graph, double click to bring up a dialog of options associated with the element, or use the toolbar or the right mouse button menus to remove the object entirely.

The rest of this section describes these operations in more detail.

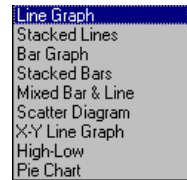
### The Graph Options Dialog

Many of the most important options for a graph are associated with the plot area of the graph. Double clicking inside the plot area brings up the Graph Options dialog:



### Graph Type

The combo box in the upper left corner of the dialog allows you to select a graph type. To change the graph type, simply select an entry from the combo box. Note that some graph types are not available for different data types (for example, you cannot view a scatter diagram for a single series). Furthermore, some views do not allow you to change the graph type. In such cases, the **Graph Type** will display **Special**, and you will not have access to the entries in the combo box.



The basic graph types are as follows:

- **Line Graph** displays a plot of the series, with each value plotted vertically against either an observation indicator or time.
- **Stacked Lines** displays results for several series, with each line representing the sum of all the preceding series. Put differently, each series value is plotted against an indicator or time as the vertical distance between each successive line.
- **Bar Graph** displays the value of each series as the height of a bar.
- **Stacked Bars** displays the value of each series as the height of a colored segment of the bar, where the overall height of each bar is the sum of all series.
- **Mixed Bar & Line** displays the first series as a bar graph and the remaining series as lines.
- **Scatter Diagram** displays a scatter with the first series on the horizontal axis and the remaining series in the vertical axis, each with a different symbol.

- **XY Line Graph** plots the first series on the horizontal axis and the remaining series in the vertical axis, each connected as a line.
- **High-Low** displays the high-low-close chart where the first and second series are connected as a vertical line and the third series (optional) is plotted as a symbol. See Chapter 8 for examples of the use of high-low-close charts.
- **Pie Chart** displays each observation as a pie with each series shown as a wedge in a different color, where the width of the wedge is proportional to the percentage contribution of the series to the sum of all series for that observation. Series with negative values are dropped from the chart.

### Graph Attributes

You may also change the **Graph Attributes** listed in the left column of the dialog box. Each attribute can be turned on or off by clicking the box to the left of the attribute. You may put a box around the graph, indent the graph within the box, display a legend identifying the series, draw vertical or horizontal grid lines, draw a horizontal line corresponding to the zero value on the vertical axis, and use color.

### Bars and Lines

These options control the appearance of lines, bars, or wedges in a pie chart. You may let EViews make the choice automatically, or you can choose to make the lines dashed, make the areas solid color, or use patterns or hatching. If you select **Auto choice**, EViews uses different colors to distinguish lines on your color monitor. When you print the graph on a monochrome printer, EViews will substitute patterns for the colors.

### Graph Scaling Options

These options let you specify a common single scale on the left applying to all series, or dual scales with a left scale for the first series and a right scale for the remaining series. **No crossing** sets up the scales so the series do not cross, while **Crossing** sets the scales to show the full variation in the series and the lines may cross in the graph.

These options control the behavior of the scales:

- **Optimized** sets the scales to show the full variation of the series with round numbers on the axis scale.
- **Force through zero** sets the scale on the vertical axis to include the value zero.
- **Logarithmic scaling** is available only for series with nonnegative values.
- **Normalize Data** rescales each series to have mean zero and standard deviation of one.

**Use Manual Scaling** allows you to specify the maximum and minimum values for each axis in the graph. Graphs potentially have three axes. For line and bar graph types, the first scale is the left axis and the second scale is the right axis (if you choose dual scales). For XY line graphs and scatter diagrams, the first scale is the horizontal axis, the second scale is the left axis, and the third scale is the right axis (if you choose dual scales).

Note that if you specify the maximum scale smaller than the largest value of the series or if you specify the minimum scale larger than the smallest value of the series in the sample, EViews will ignore the specified scale so that all observations are displayed in the graph.

### Line Graph Options

These options determine whether to show just lines connecting the data points for each series, symbols marking the data points, or both.

**Line Patterns** lets you choose the type of line used to plot each series; click in the box and a menu of options will drop down. Note that if you are plotting your series in color, selecting the line pattern will also change the color of the displayed line. In fact, if you wish to change the color of your lines, you will do so by selecting an alternative line pattern.

### Bar Graphs Options

These options allow you to put values (height) of each bar in the graph as labels and to control the spacing between bars of adjacent observations. For large samples with many bars, the bars are not labeled and have no spacing between them.

### Scatter Diagram Options

These options allows you to connect the consecutive points with lines and to plot the fitted line from a bivariate regression of the vertical axis series on the horizontal axis series.

### Pie Graph Options

The **Label** option puts the date or observation number of each pie in the graph. When there are many pies, the pies are labeled sparsely.

### Setting Fonts

You can change the fonts used in the graph labels (the axes, legends, and added text) by clicking the **Fonts** button. The Font button in the Graph Options Dialog sets the default font for all graph components. If you wish, you may then set individual fonts for axes, legends and text.

### Default Graph Options

The graph options in effect when an object is first created are determined from the global graph options. To change the default graph option settings, click **Options/Graphics Defaults...** in the main menu. The settings selected in this dialog will be applied to new objects when they are cre-

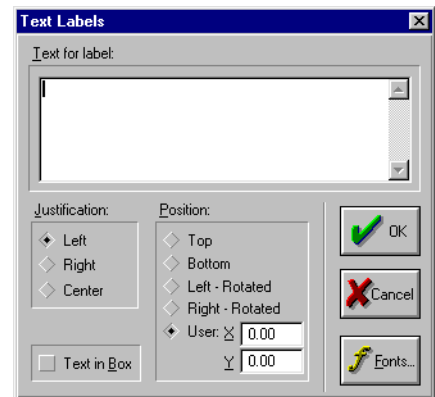
ated. Existing objects will not be affected. Changes made to a view of an object in the Graph Options dialog will modify these default settings and will generally be kept with the object. This is not the case for manual scaling, line patterns and font selection all of which will be stored only with a frozen graph object.

## Adding Text

You can customize a graph by adding one or more lines of text anywhere on the graph. This can be useful for labeling a particular observation or period, or for adding titles or remarks to the graph. In a frozen graph object, simply click on the **AddText** button in the toolbar or select **Procs/Add text....** The Text Label dialog is displayed.

Enter the text you want to display in the large edit field. Spacing and capitalization (upper and lower case letters) will be preserved. If you want to enter more than one line, press ENTER after each line.

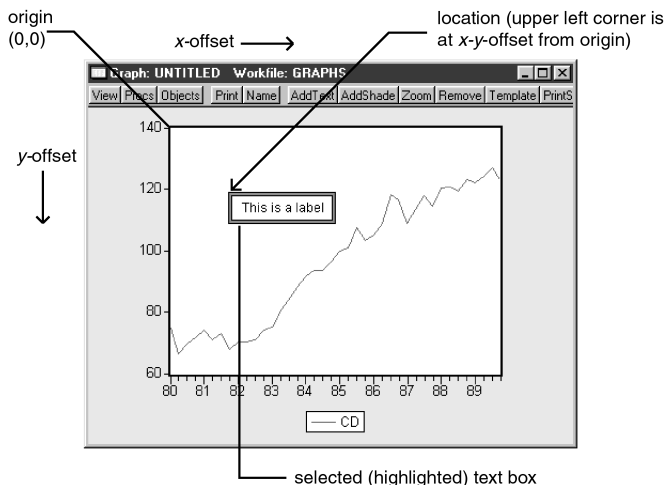
- The **Justification** options determine how multiple line labels will be aligned relative to each other.
- **Text in Box** puts the label in a rectangle box.
- **Fonts** allows you to select the fonts for the label.



## Position

These options determine the initial placement of the text within your graph. The first four options place the text at the indicated position outside the graph.

You can also place the text by specifying its coordinates. Coordinates are set in virtual inches, with the origin at the upper left-hand corner of the graph.



The X-axis position increases as you move to the right of the origin, while the Y-axis increases as you move down from the origin. All graphs are  $4 \times 3$  virtual inches except scatter diagrams, which are  $3 \times 3$  virtual inches. Thus, the  $X = 4$ ,  $Y = 3$  position refers to the lower right hand corner of the graph. Labels will be placed with the upper left-hand corner of the enclosing box at the specified coordinate.

You can change the position of text added to the graph by selecting the text box and dragging it to the position you choose.

After dragging to the desired position, you may double click on the text to bring up the Table Labels dialog and to check the coordinates of that position or to make changes to the text.

## Adding Shading

You may add a shaded region to intervals of observations or place a vertical line in a graph. From a graph object, click on the **AddShade** button in the toolbar or select **Procs/Add shading....** The Shading and Vertical Lines dialog will appear. Enter the start and end date (or observation number) of the interval you want to shade. If you enter the same date/observation number for the start and end, EViews will place a vertical line at that date.

To remove shading, click on the shaded area to select the region, and press the **Remove** button on the toolbar. **AddShade** is currently available only for graphs that have dates (or observations) along the horizontal axis.

For multiple graphs, choosing **AddShade** from the toolbar adds shading to all graphs in the specified range. If you want to shade only one graph, first highlight the graph to which you wish to add shading before clicking **AddShade**.

## Customizing Axis Labels

If you double click on the vertical or horizontal axis outside the plot area, the Scale Labels dialog will appear. This dialog allows you to choose whether the axis is marked with ticks, numbers, or both, and whether the ticks are outside the graph, inside the graph, or both. You can select the fonts for the tick numbers by clicking the **Font** button in the dialog.

## Customizing Legends

If you double click on the legend, the Legend dialog box appears. This dialog provides you the following options:

- Modify the labels of the series so that they are different from the names of the series.
- Specify the position of the legend in the graph; see Position above for an explanation of coordinates in graphs.
- To place the legend without a box, click off the check mark in front of **Legend in Box**.
- **Columns** specifies the number of columns in which to place the legends when there are multiple series in the graph.
- Press the **Font** button to change the legend font.

For scatter diagrams, you can change the appearance of the axis labels. Double click on the label and specify the options in the Text Labels dialog.

You can move the legend box at any time by dragging it with your mouse.

## Removing Graph Elements

The **Remove** button in the graph toolbar removes the selected element of a frozen graph. For example, to remove text that you have placed on the graph, click on the text. A border will appear around the text. Press the **Remove** button and it will disappear. The same method can be applied to legends, scales, and shading.

## Graph Templates

Having put a lot of effort into getting a graph to look just the way you want it, you may want to use the same options in another graph. EViews allows you to use any graph as a *template* for a new or existing graph. You can think of a template as a graph style that can be applied to other graphs.

First, name the graph object that you want to use as template. Then click the **Template** button in the toolbar of the graph object to which you wish to apply the template. Type the name of the graph object to use as a template. There are two template options:

- **Copy options** applies only the Graph Options settings of the template graph to the graph you are working on.
- **Copy options, text and shading** copies any text labels and shading in the template to the graph as well as its options.

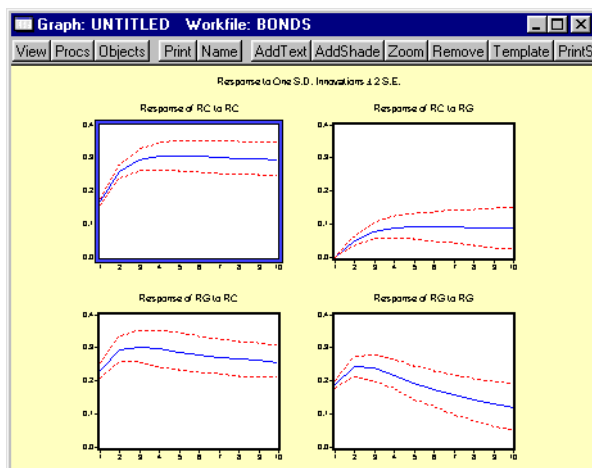
If you select **Template** from a multiple graph window, the template options will be applied to each graph in the multiple graph object.

## Graph Commands

For those of you who wish to automate some of these procedures, for example to produce a regular report, EViews allows you to perform graph customization from the command line or using programs. For details, see the *Command and Programming Reference*.

## Multiple Graph Objects

Some views are made up of multiple graphs. Like single graph views, these may be turned into graph objects by freezing. For example, the impulse response view of a VAR can display multiple graphs in a single view. Clicking on the **Freeze** button will create a graph object containing multiple graphs:



You may also create a graph object containing multiple graphs by combining existing named graph objects. Simply select the desired graphs and then double click on any one of the highlighted names.

An alternative method of combining graphs is to select **Quick/Show...** and enter the names of the graphs.

- You can position the graphs. Single click to select an individual graph (it will be highlighted, as depicted above, where the upper left-hand graph is selected), and drag it to a new position.
- You can delete a selected graph. Select the desired graph and press the **Remove** button on the graph toolbar.
- You may set options for each individual graph. To change the options for a single graph, double click on the graph area to open the graph options dialog.

## Multiple Graph Options

For multiple graphs, you can specify different options for each graph as described above, or you can apply the same options to all graphs.

To apply the same options, double click anywhere on the yellow background of the graph or select **Procs**. EViews will display a menu.



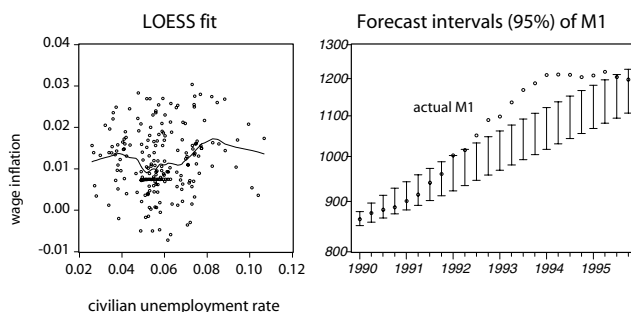
These menu items set options that apply to all graphs in the graph object.

Change options on all graphs...  
Change scaling on all graphs...  
Position and align graphs...  
Add shading to all graphs...  
Add text  
Template...

- To display all graphs with a common scale, select the **Change scaling on all graphs...** option.
- While each single graph in a multiple graph can be freely positioned by dragging the graph, you may wish to globally align graphs in columns and control the overall spacing between graphs. To globally position your graphs, elect **Procs/Position and align graphs....**
- If all of your graphs share a common observation axis (x-axis), you can add shading to each graph in the object, by selecting **Add shading to graphs...** and entering an observation pair.
- Selecting **Add text...** allows you to annotate your multiple graph. Note that adding an item to the multiple graph differs from adding it to an individual graph since it will not move with the individual graph.

## An Example

Here is an example of an annotated and customized combined graph created in EViews and copy-and-pasted into a word processor:



## Printing Graphs

Clicking on the **Print** button on the graph view or graph object window toolbar will print the graph. You can control printing options using **File/Print Setup** on the main EViews menu.

Most of the options are self-explanatory. If you wish to print your graph in color using your color printer, make certain that the **Print in color** box is checked. Conversely, if you are printing to a black and white printer, you should make certain this box is not checked so that EViews will substitute line patterns for colors.

## Printing Graphs as PostScript Files

EViews uses the standard Windows printer services for all of its printing. This means that the standard Windows methods of setting up printing to a file will work. Under Windows 95, it is very simple to set this up. From the taskbar, choose **Start/Settings/Printers**. Double click on **Add Printer**, click **Next**, click on **Local**, and select a PostScript printer. We have found that selecting the HP LaserJet 4M or the Apple LaserWriter Pro drivers (by clicking on the manufacturer, and then the printer name) appears to work well, though you may wish to experiment with different drivers depending upon your printing hardware.

Once you have selected a printer, select FILE: to instruct Windows to print to a file and then tell Windows whether you wish this to be the default printer. Windows will then create a new printer icon representing file capture of the postscript. This “printer” may be selected from within any Windows application including EViews.

## Copying Graphs to Other Windows Programs

You can incorporate an EViews graph directly into a document in your Windows word processor. To do this, first activate the object window containing the graph you wish to move; click anywhere in the window so that the titlebar changes to a bright color. Then click **Edit/Copy** on the EViews main menu; the Copy Graph as Metafile dialog box appears.

You can copy the graph to the Windows clipboard or to a disk file. You can request that the graph be in color and that its lines be in bold. We recommend that you copy graphs in black-and-white unless you will be printing to a color printer. If you plan to use the metafile in another document or application, you should select **Make metafile placeable**.

If you copy a graph to the clipboard, you can switch to your word processor and paste the graph into your document. Standard programs such as Word or WordPerfect will give you a graph which can be sized, positioned, and modified within the program. You can also paste graphs into drawing programs, and make further modifications before pasting into your word processor or other software.

## Tables

Any views that are not graphical are displayed either as tables or text windows. Tables contain formatted text that is aligned in columns and rows. Examples of views displayed in tables are the SpreadSheet views and the Estimation Output views. Just as there are two types of graphs, there are two types of tables: table views and table objects. Table objects can be created by freezing table views. As with graph objects, table objects are “not live” in the sense that they do not reflect the current contents of the underlying object, but are based instead upon the contents of the object at the time the object was frozen.

## Table Options

As with graph objects, there are editing options for tables that are more extensive than for views. These options are available from the **Procs** menu or as buttons on the table window toolbar. When you edit a table, you either highlight the cells you want to edit or refer to a rectangle specified by upper-left and lower-right cells. Cells are identified by the column letter and row number displayed in the margins of the table. If you highlight the cells before you push the editing button, the dialogs will propose that editing be applied to the highlighted cells.

- **Font** allows you to choose a font to be used in the table.
- **Insert-Delete (InsDel)** inserts or deletes rows or columns at the desired position in the table.
- **Column Width (Width)** changes the width of the columns in the table. The width is measured by the number of characters and depends on the font used.
- **Number Cell Format (Numbers)** controls the format for some or all of the numbers in the table. **Fixed characters** specifies the total number of characters to display, while **Fixed decimal** specifies the number of digits to the right of the decimal point. Note that the digits may not be visible if the cell width is not wide enough; use **Column Width** to increase the cell width. **Justification** entries set in the **Number Cell Format** dialog only apply to numbers. You should use **Justify** for justification of all selected cells.
- **Justification (Justify)** justifies all numbers and text for the selected cells.
- **Horizontal Lines (Lines)** adds or removes (double) lines from the specified cells. Note that lines will overlap with any other existing text or numbers in the cell. To place a line by itself, insert an empty row by clicking **InsDel**.
- **Grid +/–** toggles on or off the grid marking the cells in the table.
- **Title** allows you to put a header title at the top center of the table. This is different from **Name**, which provides the object name for the table in the workfile.
- **Edit +/–** toggles on or off edit mode that allows you edit text or numbers in the table.

## Table Commands

For those of you who wish to automate some of these procedures, say to produce a monthly report, EViews allows you to perform considerable customization from the command line or using programs. For details, see the *Command and Programming Reference*.

## Copying Tables to Other Windows Programs

You can also cut and paste a table into spreadsheet or word processing software. Highlight the area of the table you want to copy and then choose **Edit/Copy** from the main menu. A dialog

box will provide you with the option of copying the numbers as they appear in the table, or at the highest internal precision. After you make a choice, switch to your application and select **Edit/Paste**.

EViews copies tables as rich text format (RTF) files, allowing you to preserve the formatting information that is built into the table. Thus, if you copy-and-paste a table from EViews into Microsoft Word or another program which supports RTF, you will create a nicely formatted table containing your results:

Dependent Variable: M1				
Method: Least Squares				
Date: 09/07/97 Time: 12:51				
Sample(adjusted): 1952:2 1979:4				
Included observations: 111 after adjusting endpoints				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-4.881480	1.532169	-3.185993	0.0019
PCGDPD	-0.033838	0.202541	-0.167066	0.8676
GDPD	0.008857	0.005484	1.615040	0.1093
M1(-1)	1.014426	0.013662	74.25210	0.0000
RD	-0.432699	0.378003	-1.144697	0.2549
R-squared	0.998632	Mean dependent var		214.6029
Adjusted R-squared	0.998581	S.D. dependent var		85.66340
S.E. of regression	3.227174	Akaike info criterion		5.225090
Sum squared resid	1103.953	Schwarz criterion		5.347140
Log likelihood	-284.9925	F-statistic		19350.15
Durbin-Watson stat	2.770462	Prob(F-statistic)		0.000000

Some word processors provide the option of pasting the contents of the clipboard as unformatted files. If you wish to paste the table as unformatted text, you should select **Edit/Paste Special**.

## Text Objects

Some output views have no formatting and are simple displays of text information. Examples are representations of an equation and results from X-11 seasonal adjustment. If you freeze one of these views, you will create a text object.

You can also create a blank text object by selecting **Objects/New Object/Text** or by simply typing text in the command window. Text objects may be used whenever you wish to capture textual data that does not contain any formatting information.

## Commands

The `freeze` command freezes the specified view of the named object. Provide a name for the frozen object in parentheses after the `freeze` command. For example, to freeze the histogram view of a series named `LWAGE` to a graph object named `LW_HIST`, type

```
freeze(lw_hist) lwage.hist
```

To freeze the simple scatter diagram of the series in the group GRP1 to a graph named GRA1, type

```
freeze(gra1) grp1.scat
```

To combine two graph objects named GRA1 and GRA2 into a graph object named BIGGRA, type

```
freeze(biggra) gra1 gra2
```

See the *Command and Programming Reference* for a more complete discussion of these and other related commands.



## Part III. Basic Single Equation Analysis

---

Chapters 11–15 describe the EViews features for basic single equation analysis.

- Chapter 11. Basic Regression Models—outlines the basics of ordinary least squares estimation in EViews.
- Chapter 12. Additional Regression Models—discusses weighted least squares, two-stage least squares and nonlinear least square estimation techniques.
- Chapter 13. Time Series Regression—describes single equation regression techniques for the analysis of time series data: testing for serial correlation, estimation of ARMA and ARIMAX models, using polynomial distributed lags, and unit root tests for nonstationary time series.
- Chapter 14. Specification Testing—describes specification testing in EViews.
- Chapter 15. Forecasting—describes the fundamentals of using EViews to forecast from estimated equations.

Additional single equation techniques for autoregressive conditional heteroskedasticity, and discrete and limited dependent variable models are described in Part IV. Part V documents multiple equation analysis.





# Chapter 11. Basic Regression

---

Single equation regression is one of the most versatile and widely used statistical techniques. This chapter describes the use of basic regression techniques in EViews: specifying and estimating a regression model, performing simple diagnostic analysis, and using your estimation results in further analysis.

Subsequent chapters discuss testing and forecasting, as well as more advanced and specialized techniques such as weighted least squares, two-stage least squares (TSLS), nonlinear least squares, ARIMA/ARIMAX models, generalized method of moments (GMM), GARCH models, and qualitative and limited dependent variable models. These techniques and models all build upon the basic ideas presented in this chapter.

You will probably find it useful to own an econometrics textbook as a reference for the techniques discussed in this and subsequent chapters. Standard textbooks that we have found to be useful are listed below (in generally increasing order of difficulty):

- Pindyck and Rubinfeld (1991)
- Johnston and DiNardo (1997)
- Greene (1997)
- Davidson and MacKinnon (1993)

Where appropriate, we will also provide you with specialized references for specific topics.

## Equation Objects

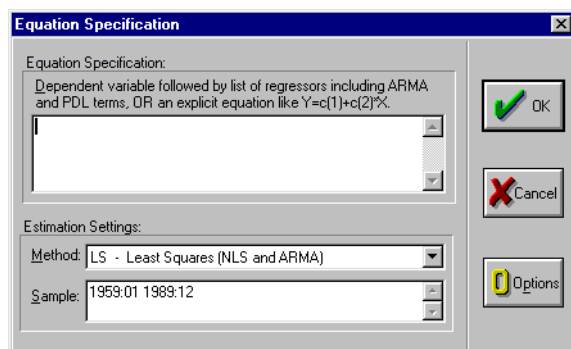
Single equation regression estimation in EViews is performed using the *equation object*. To create an equation object in EViews: select **Objects/New Object/Equation** or **Quick/Estimate Equation...** from the main menu, or simply type the keyword `equation` in the command window.

Next, you will specify your equation in the Equation Specification dialog box that appears, and select an estimation method. Below, we provide details on specifying equations in EViews. EViews will estimate the equation and display results in the equation window.

The estimation results are stored as part of the equation object so they can be accessed at any time. Simply open the object to display the summary results, or to access EViews tools for working with results from an equation object. For example, you can

retrieve the sum-of-squares from any equation, or you can use the estimated equation as part of a multi-equation model.

## Specifying an Equation in EViews



When you create an equation object, a specification dialog box is displayed.

You need to specify three things in this dialog: the equation specification, the estimation method, and the sample to be used in estimation.

In the upper edit box, you can specify the equation: the dependent (left-hand side) and independent (right-hand side) variables and the func-

tional form. There are two basic ways of specifying an equation: by list and by formula. The list-ing method is easier but may only be used with unrestricted linear specifications; the formula method is more general and can be used to specify nonlinear models or models with parametric restrictions.

### Specifying an Equation by List

The simplest way to specify a linear equation is to provide a list of variables that you wish to use in the equation. First, include the name of the dependent variable or expression, followed by a list of explanatory variables. For example, to specify a linear consumption function, CS regressed on a constant and INC, type the following in the upper field of the Equation Specification dialog:

```
cs c inc
```

Note the presence of the series name C in the list of regressors. This is a built-in EViews series that is used to specify a constant in a regression. EViews does not automatically include a constant in a regression so you must explicitly list the constant (or its equivalent) as a regressor. The internal series C does not appear in your workfile, and you may not use it outside of specifying an equation. If you need a series of ones, you can generate a new series, or use the number 1 as an auto-series.

You may have noticed that there is a pre-defined object C in your workfile. This is the default coefficient vector—when you specify an equation by listing variable names, EViews stores the estimated coefficients in this vector, in the order of appearance in the list. In the example above, the constant will be stored in C(1) and the coefficient on INC will be held in C(2).

Lagged series may be included in statistical operations using the same notation as in generating a new series with a formula—put the lag in parentheses after the name of the series. For example the specification:

```
cs cs(-1) c inc
```

tells EViews to regress CS on its own lagged value, a constant, and INC. The coefficient for lagged CS will be placed in C(1), the coefficient for the constant is C(2), and the coefficient of INC is C(3).

You can include a consecutive range of lagged series by using the word `to` between the lags. For example,

```
cs c cs(-1 to -4) inc
```

regresses CS on a constant, CS(-1), CS(-2), CS(-3), CS(-4), and INC. If you don't include the first lag, it is taken to be zero. For example,

```
cs c inc(to -2) inc(-4)
```

regresses CS on a constant, INC, INC(-1), INC(-2), and INC(-4).

You may include auto-series in the list of variables. If the auto-series expressions contain spaces they should be enclosed in parentheses. For example,

```
log(cs) c log(cs(-1)) ((inc+inc(-1)) / 2)
```

specifies a regression of the natural logarithm of CS on a constant, its own lagged value, and a two period moving average of INC.

Typing the list of series may be cumbersome, especially if you are working with many regressors. If you wish, EViews can create the specification list for you. First, highlight the dependent variable in the workfile window by single clicking on the entry. Next, CTRL-click on each of the explanatory variables to highlight them as well. When you are done selecting all of your variables, double click on any of the highlighted series, and select **Open/Equation...** The Equation Specification dialog box should appear with the names entered in the specification field. The constant C is automatically included in this list; you must delete the C if you do not wish to include the constant.

## Specifying an Equation by Formula

You will need to specify your equation using a formula when the list method is not general enough for your specification. Many, but not all, estimation methods allow you to specify your equation using a formula.

An equation formula in EViews is a mathematical expression involving regressors and coefficients. To specify an equation using a formula, simply enter the expression in the dialog in place

of the list of variables. EViews will add an implicit additive disturbance to this equation and will estimate the parameters of the model using least squares.

When you specify an equation by list, EViews converts this into an equivalent equation formula. For example, the list,

$$\log(cs) \ c \ \log(cs(-1)) \ \log(inc)$$

is interpreted by EViews as,

$$\log(cs) = c(1) + c(2)*\log(cs(-1)) + c(3)*\log(inc)$$

Equations do not have to have a dependent variable followed by an equal sign and then an expression. The = sign can be anywhere in the formula, as in:

$$\log(urate) + c(1)*dmr = c(2)$$

The residuals for this equation are given by:

$$\epsilon = \log(urate) - c(1)dmr - c(2). \quad (11.1)$$

EViews will minimize the sum-of-squares of these residuals.

If you wish, you can specify an equation as a simple expression, without a dependent variable and an equal sign. If there is no equal sign, EViews assumes that the entire expression is the disturbance term. For example, if you specify an equation as

$$c(1)*x + c(2)*y + 4*z$$

EViews will find the coefficient values that minimize the sum of squares of the given expression, in this case  $(C(1)*X + C(2)*Y + 4*Z)$ . While EViews will estimate an expression of this type, since there is no dependent variable, some regression statistics (e.g. R-squared) are not reported and the equation cannot be used for forecasting. This restriction also holds for any equation that includes coefficients to the left of the equal sign. For example, if you specify,

$$x + c(1)*y = c(2)*z$$

EViews finds the values of  $C(1)$  and  $C(2)$  that minimize the sum of squares of  $(X + C(1)*Y - C(2)*Z)$ . The estimated coefficients will be identical to those from an equation specified using:

$$x = -c(1)*y + c(2)*z$$

but some regression statistics are not reported.

The two most common motivations for specifying your equation by formula are to estimate restricted and nonlinear models. For example, suppose that you wish to constrain the coefficients on the lags on the variable  $X$  to sum to one. Solving out for the coefficient restriction leads to the following linear model with parameter restrictions:

$$y = c(1) + c(2)*x + c(3)*x(-1) + c(4)*x(-2) + (1-c(2)-c(3)-c(4))*x(-3)$$

To estimate a nonlinear model, simply enter the nonlinear formula. EViews will automatically detect the nonlinearity and estimate the model using nonlinear least squares. For details, see Chapter 12, in the section “Nonlinear Least Squares” on page 260.

You can elect to use a different coefficient vector. To create a new coefficient vector, choose **Objects/New Object...** and select **Matrix-Vector-Coef** from the main menu, type in a name for the coefficient vector, and click **OK**. In the **New Matrix** dialog box that appears, select **Coefficient Vector** and specify how many rows there should be in the vector. The object will be listed in the workfile directory with the coefficient vector icon (the little  $\alpha$ ).

You may then use this coefficient vector in your specification. For example, suppose you created coefficient vectors A and BETA, each with a single row. Then you can specify your equation using the new coefficients in place of C:

$$\log(cs) = a(1) + beta(1)*\log(cs(-1))$$

## Estimating an Equation in EViews

### Estimation Methods

Having specified your equation, you now need to choose an estimation method. Click on the **Method:** entry in the dialog and you will see a drop-down menu listing estimation methods.

Standard, single-equation regression is performed using least squares. The other methods are described in subsequent chapters.

Equations estimated by ordinary least squares and two-stage least squares, GMM, and ARCH can be specified with a formula. Nonlinear equations are not allowed with binary, ordered, censored, and count models, or in equations with ARMA terms.

LS - Least Squares (NLS and ARMA)  
 TSLS - Two-Stage Least Squares (TSNLS and ARMA)  
 ARCH - Autoregressive Conditional Heteroskedasticity  
 GMM - Generalized Method of Moments  
 BINARY - Binary choice (logit, probit, extreme value)  
 ORDERED - Ordered choice  
 CENSORED - Censored data (tobit)  
 COUNT - Integer count data

### Estimation Sample

You should also specify the sample to be used in estimation. EViews will fill out the dialog with the current workfile sample, but you can change the sample for purposes of estimation by entering your sample string or object in the edit box (see “Samples” on page 58 of Chapter 4 for details). Changing the estimation sample does not affect the current workfile sample.

If any of the series used in estimation contain missing data, EViews will temporarily adjust the estimation sample of observations to exclude those observations (listwise exclusion). EViews

notifies you that it has adjusted the sample by reporting the actual sample used in the estimation results:

```
Dependent Variable: Y
Method: Least Squares
Date: 08/19/97   Time: 10:24
Sample(adjusted): 1959:01 1989:12
Included observations: 340
Excluded observations: 32 after adjusting endpoints
```

---

Here we see the top of an equation output view. EViews reports that it has adjusted the sample. Out of the 372 observations in the period 59.01–89.12, EViews uses the 340 observations with observations for all of the relevant variables.

You should be aware that if you include lagged variables in a regression, the degree of sample adjustment will differ depending on whether data for the pre-sample period are available or not. For example, suppose you have nonmissing data for the two series M1 and IP over the period 59.01–89.12 and specify the regression as

```
m1 c ip ip(-1) ip(-2) ip(-3)
```

If you set the estimation sample to the period 59.01–89.12, EViews adjusts the sample to:

```
Dependent Variable: M1
Method: Least Squares
Date: 08/19/97   Time: 10:49
Sample: 1960:01 1989:12
Included observations: 360
```

since data for IP(-3) are not available until 59.04. However, if you set the estimation sample to the period 60.01–89.12, EViews will not make any adjustment to the sample since all values of IP(-3) are available during the estimation sample.

Some operations, most notably estimation with MA terms and ARCH, do not allow missing observations in the middle of the sample. When executing these procedures, an error message is displayed and execution is halted if an NA is encountered in the middle of the sample. EViews handles missing data at the very start or the very end of the sample range by adjusting the sample endpoints and proceeding with the estimation procedure.

## Estimation Options

EViews provides a number of estimation options. These options allow you to weight the estimating equation, to compute heteroskedasticity and auto-correlation robust covariances, and to control various features of your estimation algorithm. These options are discussed in detail in later chapters.

## Equation Output

When you click **OK** in the Equation Specification dialog, EViews displays the equation window displaying the estimation output view:

Dependent Variable: LOG(M1)				
Method: Least Squares				
Date: 08/18/97 Time: 14:02				
Sample: 1959:01 1989:12				
Included observations: 372				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1.699912	0.164954	-10.30539	0.0000
LOG(IP)	1.765866	0.043546	40.55199	0.0000
TB3	-0.011895	0.004628	-2.570016	0.0106
R-squared	0.886416	Mean dependent var	5.663717	
Adjusted R-squared	0.885800	S.D. dependent var	0.553903	
S.E. of regression	0.187183	Akaike info criterion	-0.505429	
Sum squared resid	12.92882	Schwarz criterion	-0.473825	
Log likelihood	97.00980	F-statistic	1439.848	
Durbin-Watson stat	0.008687	Prob(F-statistic)	0.000000	

Using matrix notation, the standard regression may be written as:

$$y = X\beta + \epsilon \quad (11.2)$$

where  $y$  is a  $T$ -dimensional vector containing observations on the dependent variable,  $X$  is a  $T$  by  $k$  matrix of independent variables,  $\beta$  is a  $k$ -vector of coefficients, and  $\epsilon$  is a  $T$ -vector of disturbances.  $T$  is the number of observations and  $k$  is the number of right-hand side regressors.

In the output above,  $y$  is  $\log(M1)$ ,  $X$  consists of three variables C,  $\log(IP)$ , and TB3, where  $T = 372$  and  $k = 3$ .

## Coefficient Results

### Regression Coefficients

The column labeled “Coefficient” depicts the estimated coefficients. The least squares regression coefficients  $b$  are computed by the standard OLS formula

$$b = (X'X)^{-1}X'y \quad (11.3)$$

If your equation is specified by list, the coefficients will be labeled in the “Variable” column with the name of the corresponding regressor; if your equation is specified by formula, EViews lists the actual coefficients, C(1), C(2), etc.

For the simple linear models considered here, the coefficient measures the marginal contribution of the independent variable to the dependent variable, holding all other variables fixed. If present, the coefficient of the C is the constant or intercept in the regression—it is the base level

of the prediction when all of the other independent variables are zero. The other coefficients are interpreted as the slope of the relation between the corresponding independent variable and the dependent variable, assuming all other variables do not change.

### Standard Errors

The “Std. Error” column reports the estimated standard errors of the coefficient estimates. The standard errors measure the statistical reliability of the coefficient estimates—the larger the standard errors, the more statistical noise in the estimates. If the errors are normally distributed, there are about 2 chances in 3 that the true regression coefficient lies within one standard error of the reported coefficient, and 95 chances out of 100 that it lies within two standard errors.

The covariance matrix of the estimated coefficients is computed as,

$$\text{var}(b) = s^2(X'X)^{-1}; \quad s^2 = \hat{\epsilon}'\hat{\epsilon}/(T-k); \quad \hat{\epsilon} = y - Xb \quad (11.4)$$

where  $\hat{\epsilon}$  is the residual. The standard errors of the estimated coefficients are the square roots of the diagonal elements of the coefficient covariance matrix. You can view the whole covariance matrix by choosing **View/Covariance Matrix**.

### *t*-Statistics

The *t*-statistic, which is computed as the ratio of an estimated coefficient to its standard error, is used to test the hypothesis that a coefficient is equal to zero. To interpret the *t*-statistic, you should examine the probability of observing the *t*-statistic given that the coefficient is equal to zero. This probability computation is described below.

### Probability

The last column of the output shows the probability of drawing a *t*-statistic as extreme as the one actually observed, under the assumption that the errors are normally distributed, or that the estimated coefficients are asymptotically normally distributed.

This probability is also known as the *p-value* or the *marginal significance level*. Given a *p*-value, you can tell at a glance if you reject or accept the hypothesis that the true coefficient is zero against a two-sided alternative that it differs from zero. For example, if you are performing the test at the 5% significance level, a *p*-value lower than .05 is taken as evidence to reject the null hypothesis of a zero coefficient. If you want to conduct a one-sided test, the appropriate probability is one-half that reported by EViews.

For the above example output, the hypothesis that the coefficient on TB3 is zero is rejected at the 5% significance level but not at the 1% level. However, if theory suggests that the coefficient on TB3 cannot be positive, then a one-sided test will reject the zero null hypothesis at the 1% level.

The *p*-values are computed from a *t*-distribution with  $T - k$  degrees of freedom.



## Summary Statistics

### R-squared

The R-squared ( $R^2$ ) statistic measures the success of the regression in predicting the values of the dependent variable within the sample. In standard settings,  $R^2$  may be interpreted as the fraction of the variance of the dependent variable explained by the independent variables. The statistic will equal one if the regression fits perfectly, and zero if it fits no better than the simple mean of the dependent variable. It can be negative for a number of reasons, for example, if the regression does not have an intercept or constant, if the regression contains coefficient restrictions, or if the estimation method is two-stage least squares.

EViews computes the (centered)  $R^2$  as

$$R^2 = 1 - \frac{\hat{\epsilon}'\hat{\epsilon}}{(y - \bar{y})'(y - \bar{y})}; \quad \bar{y} = \sum_{t=1}^T y_t / T \quad (11.5)$$

where  $\bar{y}$  is the mean of the dependent (left-hand) variable.

### Adjusted R-squared

One problem with using  $R^2$  as a measure of goodness of fit is that the  $R^2$  will never decrease as you add more regressors. In the extreme case, you can always obtain an  $R^2$  of one if you include as many independent regressors as there are sample observations.

The adjusted  $R^2$ , commonly denoted as  $\bar{R}^2$ , penalizes the  $R^2$  for the addition of regressors which do not contribute to the explanatory power of the model. The adjusted  $R^2$  is computed as

$$\bar{R}^2 = 1 - (1 - R^2) \frac{T-1}{T-k} \quad (11.6)$$

The  $\bar{R}^2$  is never larger than the  $R^2$ , can decrease as you add regressors, and for poorly fitting models, may be negative.

### Standard Error of the Regression (S.E. of regression)

The standard error of the regression is a summary measure based on the estimated variance of the residuals. The standard error of the regression is computed as

$$s = \sqrt{\hat{\epsilon}'\hat{\epsilon} / (T - k)} \quad (11.7)$$

### Sum of Squared Residuals

The sum of squared residuals can be used in a variety of statistical calculations, and is presented separately for your convenience:

$$\hat{\epsilon}'\hat{\epsilon} = \sum_{t=1}^T (y_t - X_t'b)^2 \quad (11.8)$$

### Log Likelihood

EViews reports the value of the log likelihood function (assuming normally distributed errors) evaluated at the estimated values of the coefficients. Likelihood ratio tests may be conducted by looking at the difference between the log likelihood values of the restricted and unrestricted versions of an equation.

The log likelihood is computed as

$$l = -\frac{T}{2}(1 + \log(2\pi) + \log(\hat{\epsilon}'\hat{\epsilon}/T)) \quad (11.9)$$

When comparing EViews output to that reported from other sources, note that EViews does not ignore constant terms.

### Durbin-Watson Statistic

The Durbin-Watson statistic measures the serial correlation in the residuals. The statistic is computed as

$$DW = \frac{\sum_{t=2}^T (\hat{\epsilon}_t - \hat{\epsilon}_{t-1})^2}{\sum_{t=1}^T \hat{\epsilon}_t^2} \quad (11.10)$$

See Johnston and DiNardo (1997, Table D.5) for a table of the significance points of the distribution of the Durbin-Watson statistic.

As a rule of thumb, if the DW is less than 2, there is evidence of positive serial correlation. The DW statistic in our output is very close to one, indicating the presence of serial correlation in the residuals. See Chapter 13 for a more extensive discussion of the Durbin-Watson statistic and the consequences of serially correlated residuals.

There are better tests for serial correlation. In Chapter 13, we discuss the  $Q$ -statistic, and the Breusch-Godfrey LM test, both of which are more general tests of serial correlation than the Durbin-Watson test.

### Mean and Standard Deviation (S.D.) of the Dependent Variable

The mean and standard deviation of  $y$  are computed using the standard formulas:

$$\bar{y} = \sum_{t=1}^T y_t / T; \quad s_y = \sqrt{\sum_{t=1}^T (y_t - \bar{y})^2 / (T-1)} \quad (11.11)$$

### Akaike Information Criterion

The Akaike Information Criterion (AIC) is computed as:

$$\text{AIC} = -2l/T + 2k/T \quad (11.12)$$

where  $l$  is the log likelihood (given by Equation (11.9) on page 240).

The AIC is often used in model selection for non-nested alternatives—smaller values of the AIC are preferred. For example, you can choose the length of a lag distribution by choosing the specification with the lowest value of the AIC. See Appendix F, page 599, for additional discussion.

### Schwarz Criterion

The Schwarz Criterion (SC) is an alternative to the AIC that imposes a larger penalty for additional coefficients:

$$\text{SC} = -2l/T + (k \log T)/T \quad (11.13)$$

### F-Statistic and Probability

The  $F$ -statistic tests the hypothesis that *all* of the slope coefficients (excluding the constant, or intercept) in a regression are zero. For ordinary least squares models, the  $F$ -statistic is computed as

$$F = \frac{R^2/(k-1)}{(1-R^2)/(T-k)} \quad (11.14)$$

Under the null hypothesis with normally distributed errors, this statistic has an  $F$ -distribution with  $k-1$  numerator degrees of freedom and  $T-k$  denominator degrees of freedom.

The  $p$ -value given just below the  $F$ -statistic, denoted **Prob(F-statistic)**, is the marginal significance level of the  $F$ -test. If the  $p$ -value is less than the significance level you are testing, say .05, you reject the null hypothesis that all slope coefficients are equal to zero. For the example above, the  $p$ -value is essentially zero, so we reject the null hypothesis that all of the regression coefficients are zero. Note that the  $F$ -test is a joint test so that even if all the  $t$ -statistics are insignificant, the  $F$ -statistic can be highly significant.

### Working With Regression Statistics

The regression statistics reported in the estimation output view are stored with the equation and are accessible through special @-functions. You can retrieve any of these statistics for further analysis by using these functions in *genr*, *scalar*, or *matrix* expressions. If a particular statistic is not computed for a given estimation method, the function will return an NA.

There are two kinds of @-functions: those that return a scalar value, and those that return matrices or vectors.

#### Keywords that return scalar values

@r2	$R$ -squared statistic
@rbar2	adjusted $R$ -squared statistic
@se	standard error of the regression
@ssr	sum of squared residuals
@dw	Durbin-Watson statistic
@f	$F$ -statistic
@logl	value of the log-likelihood function
@aic	Akaike information criterion
@sc	Schwarz information criterion
@jstat	scalar containing the $J$ -statistic (for GMM)
@regobs	number of observations in regression
@meandep	mean of the dependent variable
@sddep	standard deviation of the dependent variable
@ncoef	total number of estimated coefficients
@coefs(i)	coefficient $i$ , where $i$ is given by the order in which the coefficients appear in the representations view
@stderrs(i)	standard error for coefficient $i$
@tstats(i)	$t$ -statistic value for coefficient $i$
@cov(i, j)	covariance of coefficients $i$ and $j$

#### Keywords that return vector or matrix objects

@coefs	vector of coefficient values
@stderrs	vector of standard errors for the coefficients
@tstats	vector of ratios of coefficients to standard errors
@cov	matrix containing the coefficient covariance matrix

If you use these functions without reference to an equation object, EViews will use the default equation. For example, the command:

```
series y = @dw
```

creates a series named Y and assigns to every observation, the value of the Durbin-Watson statistic for the default equation.

We strongly recommend, however, to precede the statistic with the name of an equation object and a period. This accesses the appropriate statistic for the named equation. For example:

```
series y = eq1.@dw
```

assigns to Y the value of the Durbin-Watson for the equation EQ1.

Functions that return a vector or matrix object should be assigned to the corresponding object type. For example, you should assign the `@tstats` to a vector:

```
vector tstats = eq1.@tstats
```

and the covariance matrix to a matrix:

```
matrix mycov = eq1.@cov
```

You can also access individual elements of these statistics:

```
scalar pvalue = 1-@cnorm(@abs(eq1.@tstats(4)))
scalar var1 = eq1.@covariance(1,1)
```

For details on vectors and matrices, see the *Command and Programming Reference*.

## Working with Equations

### Views of an Equation

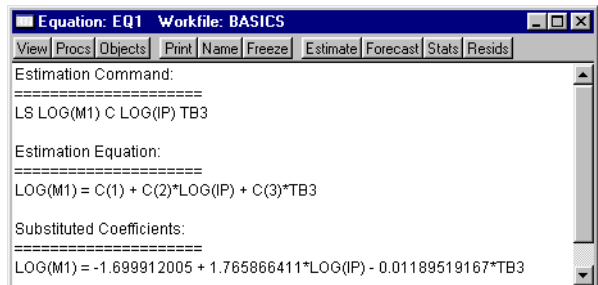
**Representations.** Displays the equation in three forms: EViews command form, as an algebraic equation with symbolic coefficients, and as an equation with the estimated values of the coefficients.

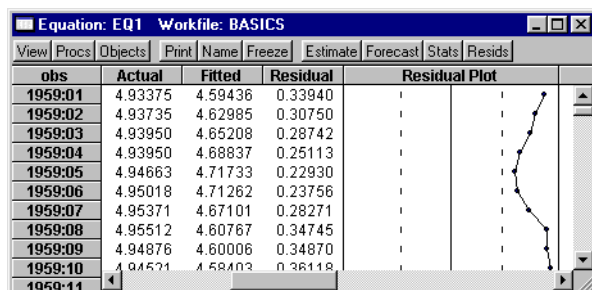
You can cut-and-paste from the representations view into any application that supports the Windows clipboard.

**Estimation Output.** Displays the equation output results described above.

#### Actual, Fitted, Residual.

These views display the actual and fitted values of the dependent variable and the residuals from the regression in tabular and graphical form. **Actual, Fitted, Residual Table** displays these values in table form.





Note that the actual value is always the sum of the fitted value and the residual. **Actual**, **Fitted**, **Residual Graph** displays a standard EViews graph of the actual values, fitted values, and residuals. **Residual Graph** plots only the residuals, while the **Standardized Residual Graph** plots the residuals divided by the estimated residual standard deviation.

**Covariance Matrix.** Displays the covariance matrix of the coefficient estimates as a spreadsheet view. To save this covariance matrix as a matrix object, use the @cov function.

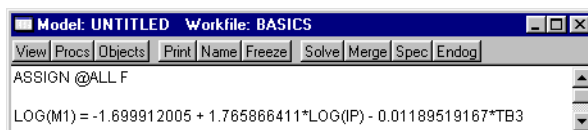
**Coefficient Tests, Residual Tests, and Stability Tests.** These are views for specification and diagnostic tests and are described in detail in Chapter 15.

## Procedures of an Equation

**Specify/Estimate...** brings up the Equation Specification dialog box so that you can modify your specification. You can edit the equation specification, or change the estimation method or estimation sample.

**Forecast...** forecasts or fits values using the estimated equation. Forecasting using equations is discussed in Chapter 14.

**Make Model** creates an untitled model containing the estimated equation:



This model can be solved in the usual manner. See Chapter 23 for information on how to use models for forecasting and simulations.

**Update Coefs from Equation** places the estimated coefficients of the equation in the coefficient vector. You can use this procedure to initialize starting values for various estimation procedures.

**Make Regressor Group** creates an untitled group comprised of all the variables used in the equation (with the exception of the constant).

**Make Residual Series...** saves the residuals from the regression as a series in the workfile. Depending on the estimation method, you may choose from three types of residuals: ordinary, standardized, and generalized. For ordinary least squares, only the ordinary residuals may be saved.

## Default Equation

Following estimation, EViews often holds the estimated coefficients and their covariance matrix, the residuals, and some summary statistics in an untitled equation object. These results are available for use in a variety of subsequent computations including the specification and diagnostic tests described in Chapter 15, and the computation of forecasts and model simulation described in Chapters 14 and 23.

Untitled equations are not saved with the workfile. You may use the **Name** button on the equation toolbar to name your equation. The equation will be saved with the workfile when the latter is saved. Once named, you can access the information in the equation at any time, even if you have just estimated several other models, or have not worked with the workfile for a long period of time.

For your convenience, EViews keeps track of a *default equation*. The default equation is the equation that is active or was the most recently active equation. The name of the default equation is shown at the upper right hand corner of the workfile window.

## Residuals from an Equation

The residuals from the default equation are stored in a series object called RESID. RESID may be used directly as if it were a regular series, except in estimation.

RESID will be overwritten whenever you estimate an equation and will contain the residuals from the latest estimated equation. To save the residuals from a particular equation for later analysis, you should save them in a different series so they are not overwritten by the next estimation command. For example, you can copy the residuals into a regular EViews series called RES1 by the command

```
series res1 = resid
```

Even if you have already overwritten the RESID series, you can always create the desired series using EViews' built-in procedures if you still have the equation object. If your equation is named EQ1, open the equation window and select **Procs/Make Residual Series**, or enter

```
eq1.makeresid res1
```

to create the desired series.

## Regression Statistics

You may refer to various regression statistics through the @-functions described above. For example, to generate a new series equal to FIT plus twice the standard error from the last regression, you can use the command

```
series plus = fit + 2*eq1.@se
```

To get the  $t$ -statistic for the second coefficient from equation EQ1, you could specify

```
eq1.@tstats(2)
```

To store the coefficient covariance matrix from EQ1 as a named symmetric matrix, you can use the command

```
sym ccov1 = eq1.@cov
```

See “Regression Statistics” on page 245 for additional details.

## Storing and Retrieving an Equation

As with other objects, equations may be stored to disk in data bank or database files. You can also fetch equations from these files.

Equations may also be copied-and-pasted to, or from, workfiles or databases.

EViews even allows you to access equations directly from your databases or another workfile. You can estimate an equation, store it in a database, and then use it to forecast in several workfiles.

See also Chapters 3 and 6 for more information about objects, databases, and object containers.

## Using Estimated Coefficients

The coefficients of an equation are listed in the representations view. By default, EViews will use the C coefficient vector when you specify an equation, but you may explicitly use other coefficient vectors in defining your equation.

These stored coefficients may be used as scalars in generating data. While there are easier ways of generating fitted values (see Chapter 14, “Forecasting from an Equation” on page 305), for purposes of illustration, note that the command:

```
series cshat = eq1.c(1) + eq1.c(2)*gdp
```

forms the fitted value of CS, CSHAT, from the OLS regression coefficients and the independent variables from the equation object EQ1.

Note that while EViews will accept a series generating equation which does not explicitly refer to a named equation:

```
series cshat = c(1) + c(2)*gdp
```

we strongly recommend that you always use named equations to identify the appropriate coefficients. In general, the coefficient vector C will contain the correct coefficient values only immediately following estimation or a coefficient update. Using a named equation, or selecting **Procs/Update coefs from equation**, guarantees that you are using the correct coefficient values.



An alternative to referring to the coefficient vector is to reference the `@coefs` elements of your equation (see p. coefficient *i*, where *i* is given by the order in which the coefficients appear in the representations view). For example, the examples above may be written as

```
series cshat=eq1.@coefs(1)+eq1.@coefs(2)*gdp
```

EViews assigns an index to each coefficient in the order that it appears in the representations view. Thus, if you estimate the equation

```
equation eq01.ls y=c(10)+b(5)*y(-1)+a(7)*inc
```

where *B* and *A* are also coefficient vectors, then

```
eq01.@coefs(1) contains C(10)
```

```
eq01.@coefs(2) contains B(5)
```

```
eq01.@coefs(3) contains A(7)
```

This method should prove useful in matching coefficients to standard errors derived from the `@stderrs` elements of the equation (see Appendix A, page 565). The `@coefs` elements allow you to refer to both the coefficients and the standard errors using a common index.

If you have used an alternative named coefficient vector in specifying your equation, you can also access the coefficient vector directly. For example, if you have used a coefficient vector named *BETA*, you can generate the fitted values by issuing the commands

```
equation eq02.ls cs=beta(1)+beta(2)*gdp
```

```
series cshat=beta(1)+beta(2)*gdp
```

where *BETA* is a coefficient vector. Again, however, we recommend that you use the `@coefs` elements to refer to the coefficients of *EQ02*. Alternatively, you can update the coefficients in *BETA* prior to use by selecting **Procs/Update coefs from equation** from the equation window. Note that EViews does not allow you to refer to the named equation coefficients `eq02.beta(1)` and `eq02.beta(2)`. You must use the expressions, `eq02.@coefs(1)` and `eq02.@coefs(2)`, instead.

## Estimation Problems

### Exact Collinearity

If the regressors are very highly collinear, EViews may encounter difficulty in computing the regression estimates. In such cases, EViews will issue an error message “Near singular matrix.” When you get this error message, you should check to see whether the regressors are *exactly* collinear. The regressors are exactly collinear if one regressor can be written as a linear combination of the other regressors. Under exact collinearity, the regressor matrix *X* does not have full column rank and the OLS estimator cannot be computed.

You should watch out for exact collinearity when you are using dummy variables in your regression. A set of mutually exclusive dummy variables and the constant term are exactly collinear. For example, suppose you have quarterly data and you try to run a regression with the specification

```
y c x @seas(1) @seas(2) @seas(3) @seas(4)
```

EViews will return a “Near singular matrix” error message since the constant and the four quarterly dummy variables are exactly collinear through the relation:

```
c = @seas(1) + @seas(2) + @seas(3) + @seas(4)
```

In this case, simply drop either the constant term or one of the dummy variables.

The textbooks listed above provide extensive discussion of the issue of collinearity.

## Commands

To declare a new equation object, follow the equation command with a name for the equation object:

```
equation eq1
```

To estimate an equation by OLS, follow the equation name with a dot and the word `ls` or `est`, the name of the dependent variable, and the names of the independent variables, each separated by a space:

```
eq1.ls cs c gdp cpi
```

regresses CS on a constant, GDP, and CPI.

Alternatively, you can specify the equation by a formula with an equal sign:

```
eq1.ls cs = c(1) + c(2)*gdp + c(3)*cpi
```

You can define and estimate an equation in one command:

```
equation eq_sale.ls sales c trend orders industry_growth
```

estimates the specified equation and stores the results in an equation named EQ\_SALE.

See the *Command and Programming Reference* for a complete list of commands and options for single equation estimation in EViews.

## Chapter 12. Additional Regression Methods

---

This chapter discusses weighted least squares, heteroskedasticity and autocorrelation consistent covariance estimation, two-stage least squares (TSLS), nonlinear least squares, and generalized method of moments (GMM). Note that most of these methods are also available in systems of equations; see Chapter 19.

Parts of this chapter refer to estimation of models which have autoregressive (AR) and moving average (MA) error terms. These concepts are discussed in greater depth in Chapter 13.

### Weighted Least Squares

Suppose that you have heteroskedasticity of known form, and that there is a series  $w$ , whose values are proportional to the *reciprocals* of the error standard deviations. You can use weighted least squares, with weight series  $w$ , to correct for the heteroskedasticity.

EViews performs weighted least squares by first dividing the weight series by its mean, then multiplying all of the data for each observation by the scaled weight series. The scaling of the weight series is a normalization that has no effect on the parameter results, but makes the weighted residuals more comparable to the unweighted residuals. The normalization does imply, however, that EViews weighted least squares is not appropriate in situations where the scale of the weight series is relevant, as in frequency weighting.

Estimation is then completed by running a regression using the weighted dependent and independent variables to minimize the sum-of-squared residuals

$$S(\beta) = \sum_t w_t^2 (y_t - x_t' \beta)^2 \quad (12.1)$$

with respect to the  $k$ -dimensional vector of parameters  $\beta$ . In matrix notation, let  $W$  be a diagonal matrix containing the scaled  $w$  along the diagonal and zeroes elsewhere, and let  $y$  and  $X$  be the usual matrices associated with the left and right-hand side variables. The weighted least squares estimator is

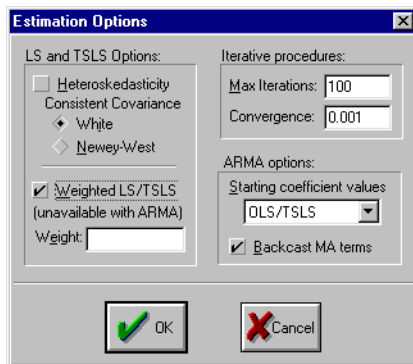
$$b_{WLS} = (X'W'WX)^{-1}X'W'y, \quad (12.2)$$

and the estimated covariance matrix is

$$\hat{\Sigma}_{WLS} = s^2(X'W'WX)^{-1}. \quad (12.3)$$

To estimate an equation using weighted least squares, first go to the main menu and select **Quick/Estimate Equation...**, then choose **LS—Least Squares (NLS and**

**ARMA**) from the combo box. Enter your equation specification and sample in the edit boxes, then push the **Options** button and click on the **Weighted LS/TSLS** option.



Fill in the blank after **Weight:** with the name of the series containing your weights, and click on **OK**. Click on **OK** again to accept the dialog and estimate the equation.

Dependent Variable: LOG(X)  
 Method: Least Squares  
 Date: 10/15/97 Time: 11:10  
 Sample(adjusted): 1891 1983  
 Included observations: 93 after adjusting endpoints  
 Weighting series: POP

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.004233	0.012745	0.332092	0.7406
LOG(X(-1))	0.099840	0.112539	0.887163	0.3774
LOG(W(-1))	0.194219	0.421005	0.461322	0.6457
<b>Weighted Statistics</b>				
R-squared	0.016252	Mean dependent var		0.009762
Adjusted R-squared	-0.005609	S.D. dependent var		0.106487
S.E. of regression	0.106785	Akaike info criterion		-1.604274
Sum squared resid	1.026272	Schwarz criterion		-1.522577
Log likelihood	77.59873	F-statistic		0.743433
Durbin-Watson stat	1.948087	Prob(F-statistic)		0.478376
<b>Unweighted Statistics</b>				
R-squared	-0.002922	Mean dependent var		0.011093
Adjusted R-squared	-0.025209	S.D. dependent var		0.121357
S.E. of regression	0.122877	Sum squared resid		1.358893
Durbin-Watson stat	2.086669			

EViews will open an output window displaying the standard coefficient results, and both weighted and unweighted summary statistics. The weighted summary statistics are based on the fitted residuals, computed using the weighted data,

$$\tilde{u}_t = w_t(y_t - x_t'b_{WLS}). \quad (12.4)$$

The unweighted results are based on the residuals computed from the original (unweighted) data,

$$u_t = y_t - x_t' b_{WLS}. \quad (12.5)$$

Following estimation, the unweighted residuals are placed in the RESID series.

If the residual variance assumptions are correct, the weighted residuals should show no evidence of heteroskedasticity. If the variance assumptions are correct, the unweighted residuals should be heteroskedastic, with the reciprocal of the standard deviation of the residual at each period  $t$  being proportional to  $w_t$ .

The weighting option will be ignored in equations containing ARMA specifications. Note also that the weighting option is not available for binary, count, censored and truncated, or ordered discrete choice models.

## Heteroskedasticity and Autocorrelation Consistent Covariances

When the form of heteroskedasticity is not known, it may not be possible to obtain efficient estimates of the parameters using weighted least squares. OLS provides consistent parameter estimates in the presence of heteroskedasticity but the usual OLS standard errors will be incorrect and should not be used for inference.

Before we describe the techniques for HAC covariance estimation, note that:

- Using the White heteroskedasticity consistent or the Newey-West HAC consistent covariance estimates does not change the point estimates of the parameters, only the estimated standard errors.
- There is nothing to keep you from combining various methods of accounting for heteroskedasticity and serial correlation. For example, weighted least squares estimation might be accompanied by White or Newey-West covariance matrix estimates.

### Heteroskedasticity Consistent Covariances (White)

White (1980) has derived a heteroskedasticity consistent covariance matrix estimator which provides correct estimates of the coefficient covariances in the presence of heteroskedasticity of unknown form. The White covariance matrix is given by:

$$\hat{\Sigma}_W = \frac{T}{T-k} (X'X)^{-1} \left( \sum_{t=1}^T u_t^2 x_t x_t' \right) (X'X)^{-1}, \quad (12.6)$$

where  $T$  is the number of observations,  $k$  is the number of regressors, and  $u_t$  is the least squares residual.

EViews provides you the option to use the White covariance estimator in place of the standard OLS formula. Open the equation dialog and specify the equation as before, then push the **Options** button. Next, click on the check box labeled **Heteroskedasticity Consistent Covariance** and click on the **White** radio button. Accept the options and click **OK** to estimate the equation.

EViews will estimate your equation and compute the variances using White's covariance estimator. You can always tell when EViews is using White covariances, since the output display will include a line to document this fact:

Dependent Variable: LOG(X)  
 Method: Least Squares  
 Date: 10/15/97 Time: 11:11  
 Sample(adjusted): 1891 1983  
 Included observations: 93 after adjusting endpoints  
 Weighting series: POP  
 White Heteroskedasticity-Consistent Standard Errors & Covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.004233	0.012519	0.338088	0.7361
LOG(X(-1))	0.099840	0.137262	0.727369	0.4689
LOG(W(-1))	0.194219	0.436644	0.444800	0.6575

### HAC Consistent Covariances (Newey-West)

The White covariance matrix described above assumes that the residuals of the estimated equation are serially uncorrelated. Newey and West (1987) have proposed a more general covariance estimator that is consistent in the presence of both heteroskedasticity and autocorrelation of unknown form. The Newey-West estimator is given by

$$\hat{\Sigma}_{NW} = \frac{T}{T-k} (X'X)^{-1} \hat{\Omega} (X'X)^{-1}, \quad (12.7)$$

where

$$\begin{aligned} \hat{\Omega} = & \frac{T}{T-k} \left\{ \sum_{t=1}^T u_t^2 x_t x_t' \right. \\ & \left. + \sum_{v=1}^q \left( \left( 1 - \frac{v}{q+1} \right) \sum_{t=v+1}^T (x_t u_t u_{t-v} x_{t-v}' + x_{t-v} u_{t-v} u_t x_t') \right) \right\} \end{aligned} \quad (12.8)$$

and  $q$ , the truncation lag, is a parameter representing the number of autocorrelations used in evaluating the dynamics of the OLS residuals  $u_t$ . Following the suggestion of Newey and West, EViews sets  $q$  to

$$q = \text{floor}(4(T/100)^{2/9}). \quad (12.9)$$

To use the Newey-West method, push the **Options** button in the estimation dialog box. Check the box labeled **Heteroskedasticity Consistent Covariance** and press the **Newey-West** radio button.

## Two-Stage Least Squares

A fundamental assumption of regression analysis is that the right-hand side variables are uncorrelated with the disturbance term. If this assumption is violated, both OLS and weighted LS are biased and inconsistent.

There are a number of situations where some of the right-hand side variables are correlated with disturbances. Some classic examples occur when:

- There are endogenously determined variables on the right-hand side of the equation.
- Right-hand side variables are measured with error.

For simplicity, we will refer to variables that are correlated with the residuals as *endogenous*, and variables that are not correlated with the residuals as *exogenous* or *predetermined*.

The standard approach in cases where right-hand side variables are correlated with the residuals is to estimate the equation using *instrumental variables* regression. The idea behind instrumental variables is to find a set of variables, termed *instruments*, that are both (1) correlated with the explanatory variables in the equation, and (2) uncorrelated with the disturbances. These instruments are used to eliminate the correlation between right-hand side variables and the disturbances.

Two-stage least squares (TSLS) is a special case of instrumental variables regression. As the name suggests, there are two distinct stages in two-stage least squares. In the first stage, TSLS finds the portions of the endogenous and exogenous variables that can be attributed to the instruments. This stage involves estimating an OLS regression of each variable in the model on the set of instruments. The second stage is a regression of the original equation, with all of the variables replaced by the fitted values from the first-stage regressions. The coefficients of this regression are the TSLS estimates.

You need not worry about the separate stages of TSLS since EViews will estimate both stages simultaneously using instrumental variables techniques. More formally, let  $Z$  be the matrix of instruments, and  $y$  and  $X$  be the dependent and explanatory variables. Then the coefficients computed in two-stage least squares are given by,

$$b_{TSLS} = (X'Z(Z'Z)^{-1}Z'X)^{-1}X'Z(Z'Z)^{-1}Z'y, \quad (12.10)$$

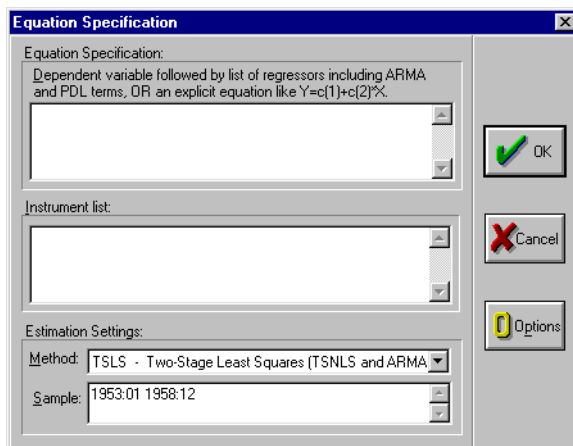
and the estimated covariance matrix of these coefficients is given by

$$\hat{\Sigma}_{TSLS} = s^2(X'Z(Z'Z)^{-1}Z'X)^{-1}, \quad (12.11)$$

where  $s^2$  is the estimated residual variance (square of the standard error of the regression).

## Estimating TSLS in EViews

To use two-stage least squares, open the equation specification box by choosing **Object/New Object/Equation...** or **Quick/Estimate Equation....** Choose **TSLS** from the **Method:** combo box and the dialog will change to include an edit window where you will list the instruments. In the edit boxes, specify your dependent variable and independent variables and the list of instruments.



There are a few things to keep in mind as you enter your instruments:

- In order to calculate TSLS estimates, your specification must satisfy the *order condition* for identification, which says that there must be at least as many instruments as there are coefficients in your equation. There is an additional rank condition which must also be satisfied. See Davidson and MacKinnon (1994) and Johnston and DiNardo (1997) for additional discussion.
- For econometric reasons that we will not pursue here, any right-hand side variables that are not correlated with the disturbances can be used as instruments.
- The constant, C, is always a suitable instrument, so EViews will add it to the instrument list if you omit it.

For example, suppose you are interested in estimating a consumption equation relating consumption (CONS) to gross domestic product (GDP), lagged consumption (CONS(-1)), a trend variable (TIME) and a constant (C). GDP is endogenous and therefore correlated with the residuals. You may, however, believe that government expenditures (G), the log of the money supply (LM), lagged consumption, TIME, and C, are exogenous and uncorrelated with the disturbances, so that these variables may be used as instruments. Your equation specification is then,

```
cons c gdp cons(-1) time
```



and the instrument list is,

```
c gov cons(-1) time lm
```

This specification satisfies the order condition for identification, which requires that there are at least as many instruments (five) as there are coefficients (four) in the equation specification.

Furthermore, all of the variables in the consumption equation that are believed to be uncorrelated with the disturbances, (CONS(-1), TIME, and C), appear both in the equation specification and in the instrument list. Note that listing C as an instrument is redundant, since EViews automatically adds it to the instrument list.

## Output from TSLS

Below, we present TSLS estimates from a regression of LOG(CS) on a constant and LOG(GDP), with the instrument list C LOG(CS(-1)) LOG(GDP(-1)):

Dependent Variable: LOG(CS)				
Method: Two-Stage Least Squares				
Date: 10/15/97 Time: 11:32				
Sample(adjusted): 1947:2 1995:1				
Included observations: 192 after adjusting endpoints				
Instrument list: C LOG(CS(-1)) LOG(GDP(-1))				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1.209268	0.039151	-30.88699	0.0000
LOG(GDP)	1.094339	0.004924	222.2597	0.0000
R-squared	0.996168	Mean dependent var	7.480286	
Adjusted R-squared	0.996148	S.D. dependent var	0.462990	
S.E. of regression	0.028735	Sum squared resid	0.156888	
F-statistic	49399.36	Durbin-Watson stat	0.102639	
Prob(F-statistic)	0.000000			

EViews identifies the estimation procedure, as well as the list of instruments in the header. This information is followed by the usual coefficient,  $t$ -statistics, and asymptotic  $p$ -values.

The summary statistics reported at the bottom of the table are computed using the formulas outlined in Chapter 11. Bear in mind that all reported statistics are only asymptotically valid. For a discussion of the finite sample properties of TSLS, see Johnston and DiNardo (1997, pp. 355–358) or Davidson and MacKinnon (1984, pp. 221–224).

EViews uses the *structural residuals*  $u_t = y_t - x_t' b_{TSLS}$  in calculating all of the summary statistics. For example, the standard error of the regression used in the asymptotic covariance calculation is computed as

$$s^2 = \sum_t u_t^2 / (T - k). \quad (12.12)$$

These structural residuals should be distinguished from the *second stage residuals* that you would obtain from the second stage regression if you actually computed the two-stage least

squares estimates in two separate stages. The second stage residuals are given by  $\hat{u}_t = \hat{y}_t - \hat{x}_t' b_{TSLs}$ , where the  $\hat{y}_t$  and  $\hat{x}_t$  are the fitted values from the first-stage regressions.

We caution you that some of the reported statistics should be interpreted with care. For example, since different equation specifications will have different instrument lists, the reported  $R^2$  for TSLs can be negative even when there is a constant in the equation.

## Weighted TSLs

You can combine TSLs with weighted regression. Simply enter your TSLs specification as above, then press the **Options** button, select the **Weighted LS/TSLs** option and enter the weighting series.

Weighted two-stage least squares is performed by multiplying all of the data, including the instruments, by the weight variable, and estimating TSLs on the transformed model. Equivalently, EViews then estimates the coefficients using the formula,

$$b_{WTSLs} = (X'W'WZ(Z'W'WZ)^{-1}Z'W'WX)^{-1} \cdot X'W'WZ(Z'W'WZ)^{-1}Z'W'Wy \quad (12.13)$$

The estimated covariance matrix is

$$\hat{\Sigma}_{WTSLs} = s^2(X'W'WZ(Z'W'WZ)^{-1}Z'W'WX)^{-1}. \quad (12.14)$$

## TSLs with AR errors

You can adjust your TSLs estimates to account for serial correlation by adding AR terms to your equation specification. EViews will automatically transform the model to a nonlinear least squares problem, and estimate the model using instrumental variables. Details of this procedure may be found in Fair (1984, pp. 210–214). The output from TSLs with an AR(1) specification looks as follows:

Dependent Variable: LOG(CS)  
 Method: Two-Stage Least Squares  
 Date: 10/15/97 Time: 11:42  
 Sample(adjusted): 1947:2 1995:1  
 Included observations: 192 after adjusting endpoints  
 Convergence achieved after 4 iterations  
 Instrument list: C LOG(CS(-1)) LOG(GDP(-1))

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1.420705	0.203266	-6.989390	0.0000
LOG(GDP)	1.119858	0.025116	44.58782	0.0000
AR(1)	0.930900	0.022267	41.80595	0.0000
R-squared	0.999611	Mean dependent var	7.480286	
Adjusted R-squared	0.999607	S.D. dependent var	0.462990	
S.E. of regression	0.009175	Sum squared resid	0.015909	
F-statistic	243139.7	Durbin-Watson stat	1.931027	
Prob(F-statistic)	0.000000			
Inverted AR Roots	.93			

The **Options** button in the estimation box may be used to change the iteration limit and convergence criterion for the nonlinear instrumental variables procedure.

### First-order AR errors

Suppose your specification is:

$$\begin{aligned} y_t &= x_t' \beta + w_t \gamma + u_t \\ u_t &= \rho_1 u_{t-1} + \epsilon_t \end{aligned} \quad (12.15)$$

where  $x_t$  is a vector of endogenous variables, and  $w_t$  is a vector of predetermined variables, which, in this context, may include lags of the dependent variable.  $z_t$  is a vector of instrumental variables not in  $w_t$  that is large enough to identify the parameters of the model.

In this setting, there are important technical issues to be raised in connection with the choice of instruments. In a widely quoted result, Fair (1970) shows that *if the model is estimated using an iterative Cochrane-Orcutt procedure*, all of the lagged left and right-hand side variables  $(y_{t-1}, x_{t-1}, w_{t-1})$  must be included in the instrument list to obtain consistent estimates. In this case, then the instrument list should include

$$(w_t, z_t, y_{t-1}, x_{t-1}, w_{t-1}). \quad (12.16)$$

EVIEWS uses an alternative, nonlinear least squares procedure described in Fair (1984). As a result, you do not need to include the lagged left and right-hand side variables in your instrument list. These variables are, however, obvious instruments, and in most cases should be included in the instrument list. There are important exceptions to this guideline, one of which is illustrated in the next section.

### Higher Order AR errors

The AR(1) result extends naturally to specifications involving higher order serial correlation. For example, if you include a single AR(4) term in your model, the natural instrument list will be

$$(w_t, z_t, y_{t-4}, x_{t-4}, w_{t-4}) \quad (12.17)$$

If you include AR terms from 1 through 4, one possible instrument list is

$$(w_t, z_t, y_{t-1}, \dots, y_{t-4}, x_{t-1}, \dots, x_{t-4}, w_{t-1}, \dots, w_{t-4}) \quad (12.18)$$

While theoretically valid, this instrument list has a large number of overidentifying instruments, which may lead to computational difficulties and large finite sample biases (Fair (1984, p. 214), Davidson and MacKinnon (1993, pp. 222-224)). In theory, adding instruments should always improve your estimates, but as a practical matter this may not be so in small samples. Thus, as you add additional AR terms to your specification, it may not be sensible to add all of the corresponding lagged endogenous and predetermined variables to the instrument list.

If you choose to omit the lagged left and right-hand side terms from the instrument list, EViews will use a hybrid instrumental variables specification; each of the lagged terms will be used as its own instrument, while the fitted values for the endogenous variables will be formed from a first-stage regression on the instruments in your list.

### Examples

Suppose that you wish to estimate the consumption function by two-stage least squares, allowing for first-order serial correlation. You may then use two-stage least squares with the variable list,

```
cons c gdp ar(1)
```

and instrument list,

```
c gov log(m1) time cons(-1) gdp(-1)
```

Notice that the lags of both the dependent and endogenous variables (CONS(-1) and GDP(-1)), are included in the instrument list.

Similarly, consider the consumption function,

```
cons c cons(-1) gdp ar(1)
```

A valid instrument list is given by

```
c gov log(m1) time cons(-1) cons(-2) gdp(-1)
```

Here we treat the lagged left and right-hand side variables from the original specification as predetermined and add the lagged values to the instrument list.

Lastly, consider the specification,

```
cons c gdp ar(1) ar(2) ar(3) ar(4)
```

If we were to include all “recommended” instruments in the list, we would have

```
c gov log(m1) time cons(-1) cons(-2) cons(-3) cons(-4) gdp(-1)
gdp(-2) gdp(-3) gdp(-4)
```

An alternative list (possibly superior in small samples) is given by,

```
c gov log(m1) time
```

The only difference that results from using the latter instrument list is that we use C, GOV, LOG(M1) and TIME in fitting CONS; the consumption and GDP lags are not used.

## TSLS with MA errors

You can also estimate two-stage least squares variable problems with MA error terms of various orders. To account for the presence of MA errors, simply add the appropriate terms to your specification prior to estimation.

## Illustration

Suppose that you wish to estimate the consumption function by two-stage least squares, accounting for first-order moving average errors. You may then use two-stage least squares with the variable list,

```
cons c gdp ma(1)
```

and instrument list,

```
c gov log(m1) time
```

In principle, you could add any lag of CONS and GDP to your instrument list, but these instruments are sufficient to satisfy the order condition.

## Technical Details

Most of the technical details are identical to those outlined above for AR errors. EVIEWS transforms the model that is nonlinear in parameters (employing backcasting, if appropriate) and then estimates the model using nonlinear instrumental variables techniques.

As above, all of the lagged endogenous and predetermined variables are candidate instruments, but since the number of lags in an MA is effectively unbounded, you obviously cannot include all of these variables in your instrument list. You should instead balance the theoretical increases in efficiency from using large numbers of instruments against the potential for computational difficulties and finite sample bias.

## Nonlinear Least Squares

Suppose that we have the regression specification

$$y_t = f(x_t, \beta) + \epsilon_t, \quad (12.19)$$

where  $f$  is a general function of the explanatory variables  $x_t$  and the parameters  $\beta$ . Least squares estimation chooses the parameter values that minimize the sum of squared residuals:

$$S(\beta) = \sum_t (y_t - f(x_t, \beta))^2 = (y - f(X, \beta))'(y - f(X, \beta)) \quad (12.20)$$

We say that a model is *linear in parameters* if the derivatives of  $f$  with respect to the parameters do not depend upon  $\beta$ ; if the derivatives are functions of  $\beta$ , we say that the model is *nonlinear in parameters*.

For example, consider the model given by

$$y_t = \beta_1 + \beta_2 \log L_t + \beta_3 \log K_t + \epsilon_t. \quad (12.21)$$

it is easy to see that this model is linear in its parameters, implying that it can be estimated using ordinary least squares.

In contrast, the equation specification

$$y_t = \beta_1 L_t^{\beta_2} K_t^{\beta_3} + \epsilon_t \quad (12.22)$$

has derivatives that depend upon the elements of  $\beta$ . There is no way to rearrange the terms in this model so that ordinary least squares can be used to minimize the sum-of-squared residuals. We must use nonlinear least squares techniques to estimate the parameters of the model.

Nonlinear least squares minimizes the sum-of-squared residuals with respect to the choice of parameters  $\beta$ . While there is no closed form solution for the parameter estimates, the estimates satisfy the first-order conditions:

$$(G(\beta))'(y - f(X, \beta)) = 0, \quad (12.23)$$

where  $G(\beta)$  is the matrix of first derivatives of  $f(X, \beta)$  with respect to  $\beta$  (to simplify notation we suppress the dependence of  $G$  upon  $X$ ). The estimated covariance matrix is given by

$$\hat{\Sigma}_{NLLS} = s^2 (G(b_{NLLS})' G(b_{NLLS}))^{-1}. \quad (12.24)$$

where  $b_{NLLS}$  are the estimated parameters. For additional discussion of nonlinear estimation, see Pindyck and Rubinfeld (1991, pp. 231-245) or Davidson and MacKinnon (1993).

### Estimating NLS Models in EViews

It is easy to tell EViews that you wish to estimate the parameters of a model using nonlinear least squares. EViews automatically applies nonlinear least squares to any regression equation that is

nonlinear in its coefficients. Simply select **Object/New Object/Equation**, enter the equation in the equation specification dialog box and click **OK**. EViews will do all of the work of estimating your model using an iterative algorithm.

A full technical discussion of iterative estimation procedures is provided in Appendix E.

### Specifying Nonlinear Least Squares

For nonlinear regression models, you will have to enter your specification in equation form using EViews expressions that contain direct references to coefficients. You may use elements of the default coefficient vector  $C$  (e.g.  $C(1)$ ,  $C(2)$ ,  $C(34)$ ,  $C(87)$ ), or you can define and use other coefficient vectors. For example,

$$y = c(1) + c(2) * (k^{c(3)} + l^{c(4)})$$

is a nonlinear specification that uses the first through the fourth elements of the default coefficient vector,  $C$ .

To create a new coefficient vector, select **Objects/New Object/Matrix-Vector-Coef/Coefficient Vector** in the main menu and provide a name. You may now use this coefficient vector in your specification. For example, if you create a coefficient vector named  $CF$ , you can rewrite the specification above as

$$y = cf(11) + cf(12) * (k^{cf(13)} + l^{cf(14)})$$

which uses the eleventh through the fourteenth elements of  $CF$ .

You can also use multiple coefficient vectors in your specification:

$$y = c(11) + c(12) * (k^{cf(1)} + l^{cf(2)})$$

which uses both  $C$  and  $CF$  in the specification.

It is worth noting that EViews implicitly adds an additive disturbance to your specification. For example, the input

$$y = (c(1) * x + c(2) * z + 4)^2$$

is interpreted as  $y_t = (c(1)x_t + c(2)z_t + 4)^2 + \epsilon_t$ , and EViews will minimize

$$S(c(1), c(2)) = \sum_t (y_t - (c(1)x_t + c(2)z_t + 4)^2)^2. \quad (12.25)$$

If you wish, the equation specification may be given by a simple expression that does not include a dependent variable. For example, the input

$$(c(1) * x + c(2) * z + 4)^2$$

is interpreted by EViews as  $-(c(1)x_t + c(2)z_t + 4)^2 = \epsilon_t$ , and EViews will minimize

$$S(c(1), c(2)) = \sum_t (-(c(1)x_t + c(2)z_t + 4)^2)^2. \quad (12.26)$$

While EViews will estimate the parameters of this last specification, the equation cannot be used for forecasting and cannot be included in a model. This restriction also holds for any equation that includes coefficients to the left of the equal sign. For example, if you specify

$$x + c(1) * y = z^{c(2)}$$

EViews will find the values of C(1) and C(2) that minimize the sum of squares of the implicit equation

$$x_t + c(1)y_t - z_t^{c(2)} = \epsilon_t, \quad (12.27)$$

but the estimated equation cannot be used in forecasting or included in a model, since there is no dependent variable.

## Estimation Options

### *Starting Values*

Iterative estimation procedures require starting values for the coefficients of the model. There are no general rules for selecting starting values for parameters. The closer to the true values the better, so if you have reasonable guesses for parameter values, these can be useful. In some cases, you can obtain good starting values by estimating a restricted version of the model using least squares. In general, however, you will have to experiment in order to find starting values.

EViews uses the values in the coefficient vector at the time you begin the estimation procedure as starting values for the iterative procedure. It is easy to examine and change these coefficient starting values.

To see the starting values, double click on the coefficient vector in the workfile directory. If the values appear to be reasonable, you can close the window and proceed with estimating your model.

If you wish to change the starting values, first make certain that the spreadsheet view of your coefficients is in edit mode, then enter the coefficient values. When you are finished setting the initial values, close the coefficient vector window and estimate your model.

You may also set starting coefficient values from the command window using the PARAM command. Simply enter the PARAM keyword, following by each coefficient and desired value:

```
param c(1) 153 c(2) .68 c(3) .15
```

sets C(1) = 153, C(2) = .68, and C(3) = .15.



See Appendix E, “Estimation Algorithms and Options”, page 591, for further details.

### *Iteration and Convergence Options*

You can control the iterative process by specifying convergence criterion and the maximum number of iterations. Press the **Options** button in the equation dialog box and enter the desired values.

EViews reports that the estimation procedure has converged if the maximum change in the coefficients (scaled by the corresponding coefficient value) is below the threshold. For example, if you set this value to .001, EViews will test for convergence by determining whether the maximum of the percentage change in the coefficients is smaller than .1 percent.

In most cases, you will not need to change the maximum number of iterations. However, for some difficult to estimate models, the iterative procedure will not converge within the maximum number of iterations. If your model does not converge within the allotted number of iterations, simply click on the **Estimate** button, and, if desired, increase the maximum number of iterations. Click on **OK** to accept the options, and click on **OK** to begin estimation. EViews will start estimation using the last set of parameter values as starting values.

These options may also be set from the global options dialog. See Appendix B, “Estimation Defaults” on page 580.

### **Output from NLLS**

Once your model has been estimated, EViews displays an equation output screen showing the results of the nonlinear least squares procedure. Below is the output from a regression of LOG(CS) on C, and the Box-Cox transform of GDP:

Dependent Variable: LOG(CS)				
Method: Least Squares				
Date: 10/15/97 Time: 11:51				
Sample(adjusted): 1947:1 1995:1				
Included observations: 193 after adjusting endpoints				
Convergence achieved after 80 iterations				
LOG(CS)= C(1)+C(2)*(GDP^C(3)-1)/C(3)				
	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	2.851780	0.279033	10.22024	0.0000
C(2)	0.257592	0.041147	6.260254	0.0000
C(3)	0.182959	0.020201	9.056824	0.0000
R-squared	0.997252	Mean dependent var		7.476058
Adjusted R-squared	0.997223	S.D. dependent var		0.465503
S.E. of regression	0.024532	Akaike info criterion		-4.562220
Sum squared resid	0.114350	Schwarz criterion		-4.511505
Log likelihood	443.2542	F-statistic		34469.84
Durbin-Watson stat	0.134628	Prob(F-statistic)		0.000000

If the estimation procedure has converged, EViews will report this fact, along with the number of iterations that were required. If the iterative procedure did not converge, EViews will report “Convergence not achieved after” followed by the number of iterations attempted.

Below the line describing convergence, EViews will repeat the nonlinear specification so that you can easily interpret the estimated coefficients of your model.

EViews provides you with all of the usual summary statistics for regression models. Provided that your model has converged, the standard statistical results and tests are *asymptotically* valid.

## Weighted NLS

Weights can be used in nonlinear estimation in a manner analogous to weighted linear least squares. To estimate an equation using weighted nonlinear least squares, enter your specification, press the **Options** button and click on the **Weighted LS/TSLS** option. Fill in the blank after **Weight:** with the name of the weight series and then estimate the equation.

EViews minimizes the sum of the weighted squared residuals:

$$S(\beta) = \sum_t w_t^2 (y_t - f(x_t, \beta))^2 = (y - f(X, \beta))' W' W (y - f(X, \beta)) \quad (12.28)$$

with respect to the parameters  $\beta$ , where  $w_t$  are the values of the weight series and  $W$  is the matrix of weights. The first-order conditions are given by

$$(G(\beta))' W' W (y - f(X, \beta)) = 0 \quad (12.29)$$

and the covariance estimate is computed as

$$\hat{\Sigma}_{WNLS} = s^2 (G(b_{WNLS})' W' W G(b_{WNLS}))^{-1}. \quad (12.30)$$

## NLS with AR errors

EViews will estimate nonlinear regression models with autoregressive error terms. Simply select **Objects/New Object/Equation...** or **Quick/Estimate Equation...** and specify your model using EViews expressions, followed by an additive term describing the AR correction enclosed in square brackets. The AR term should consist of a coefficient assignment for each AR term, separated by commas. For example, if you wish to estimate

$$\begin{aligned} CS_t &= c_1 + GDP_t^{c_2} + u_t \\ u_t &= c_3 u_{t-1} + c_4 u_{t-2} + \epsilon_t \end{aligned} \quad (12.31)$$

you should enter the specification

$$cs = c(1) + gdp^{c(2)} + [ar(1)=c(3), ar(2)=c(4)]$$

See Chapter 13, “How EViews Estimates AR Models” on page 280 for additional details.

EViews does not currently estimate nonlinear models with MA errors, nor does it estimate weighted models with AR terms—if you add AR terms to a weighted nonlinear model, the weighting series will be ignored.

## Nonlinear TSLS

Nonlinear two-stage least squares refers to an instrumental variables procedure for estimating nonlinear regression models involving functions of endogenous and exogenous variables and parameters. Suppose we have the usual nonlinear regression model:

$$y_t = f(x_t, \beta) + \epsilon_t, \quad (12.32)$$

where  $\beta$  is a  $k$ -dimensional vector of parameters, and  $x_t$  contains both exogenous and endogenous variables. In matrix form, if we have  $m \geq k$  instruments  $z_t$ , nonlinear two-stage least squares minimizes

$$S(\beta) = (y - f(X, \beta))' Z(Z'Z)^{-1} Z'(y - f(X, \beta)) \quad (12.33)$$

with respect to the choice of  $\beta$ .

While there is no closed form solution for the parameter estimates, the parameter estimates satisfy the first-order conditions:

$$G(\beta)' Z(Z'Z)^{-1} Z'(y - f(X, \beta)) = 0 \quad (12.34)$$

with estimated covariance given by

$$\hat{\Sigma}_{TSNLLS} = s^2 (G(b_{TSNLLS})' Z(Z'Z)^{-1} Z' G(b_{TSNLLS}))^{-1}. \quad (12.35)$$

## How to Estimate Nonlinear TSLS in EViews

EViews performs the estimation procedure in a single step so that you don't have to perform the separate stages yourself. Simply select **Object/New Object/Equation...** or **Quick/Estimate Equation...** Choose **TSLS** from the **Method:** combo box, enter your nonlinear specification and the list of instruments. Click **OK**.

With nonlinear two-stage least squares estimation, you have a great deal of flexibility with your choice of instruments. Intuitively you want instruments that are correlated with  $G(\beta)$ . Since  $G$  is nonlinear, you may begin to think about using more than just the exogenous and predetermined variables as instruments. Various nonlinear functions of these variables, for example, cross-products and powers, may also be valid instruments. One should be aware, however, of the possible finite sample biases resulting from using too many instruments.

## Weighted Nonlinear Two-stage Least Squares

Weights can be used in nonlinear two-stage least squares estimation. Simply add weighting to your nonlinear TSLS specification above by pressing the **Options** button, selecting **Weighted LS/TSLS** option, and entering the name of the weight series.

The objective function for weighted TSLS is,

$$S(\beta) = (y - f(X, \beta))'W'WZ(Z'W'WZ)^{-1}Z'W'W(y - f(X, \beta)). \quad (12.36)$$

The reported standard errors are based on the covariance matrix estimate given by

$$\hat{\Sigma}_{WTSNLLS} = s^2(G(b)'W'WZ(Z'W'WZ)^{-1}Z'W'WG(b))^{-1} \quad (12.37)$$

where  $b \equiv b_{WTSNLLS}$ . Note that if you add AR or MA terms to a weighted specification, the weighting series will be ignored.

## Nonlinear Two-stage Least Squares with AR errors

While we will not go into much detail here, note that EViews can estimate non-linear TSLS models where there are autoregressive error terms. EViews does not currently estimate nonlinear models with MA errors.

To estimate your model, simply open your equation specification window, and enter your nonlinear specification, including all AR terms, and provide your instrument list. For example, you could enter the regression specification

$$cs = \exp(c(1) + gdp^c(2)) + [ar(1)=c(3)]$$

with the instrument list

$$c \text{ gov}$$

EViews will transform the nonlinear regression model as described in Chapter 9, and then estimate nonlinear TSLS on the transformed specification using the instruments C and GOV. For nonlinear models with AR errors, EViews uses a Gauss-Newton algorithm. See Appendix E, page 591, for further details.

## Solving Estimation Problems

EViews may not be able to estimate your nonlinear equation on the first attempt. Sometimes, the nonlinear least squares procedure will stop immediately. Other times, EViews may stop estimation after several iterations without achieving convergence. EViews might even report that it cannot improve the sums-of-squares. While there are no specific rules on how to proceed if you encounter these estimation problems, there are a few general areas you might want to examine.

### Starting Values

If you experience problems with the very first iteration of a nonlinear procedure, the problem is almost certainly related to starting values. See the discussion above for how to examine and change your starting values.

### Model Identification

If EViews goes through a number of iterations and then reports that it encounters a “Near Singular Matrix”, you should check to make certain that your model is identified. Models are said to be non-identified if there are multiple sets of coefficients which identically yield the minimized sum-of-squares value. If this condition holds, it is impossible to choose between the coefficients on the basis of the minimum sum-of-squares criterion.

For example, the nonlinear specification,

$$y_t = \beta_1\beta_2 + \beta_2^2x_t + \epsilon_t \quad (12.38)$$

is not identified, since any coefficient pair  $(\beta_1, \beta_2)$  is indistinguishable from the pair  $(-\beta_1, -\beta_2)$  in terms of the sum-of-squared residuals.

For a thorough discussion of identification of nonlinear least squares models, see Davidson and MacKinnon (1993, Sections 2.3, 5.2 and 6.3).

### Convergence Criterion

EViews may report that it is unable to improve the sums-of-squares. This result may be evidence of non-identification or model misspecification. Alternatively, it may be the result of setting your convergence criterion too low, which can occur if your nonlinear specification is particularly complex.

If you wish to change the convergence criterion, enter the new value in the Options dialog. Be aware that increasing this value increases the possibility that you will stop at a local minimum, and may hide misspecification or non-identification of your model.

See Appendix E, “Estimation Algorithms and Options”, page 591, for further details.

## Generalized Method of Moments (GMM)

The starting point of GMM estimation is a theoretical relation that the parameters should satisfy. The idea is to choose the parameter estimates so that the theoretical relation is satisfied as “closely” as possible. The theoretical relation is replaced by its sample counterpart and the estimates are chosen to minimize the weighted distance between the theoretical and actual values. GMM is a robust estimator in that, unlike maximum likelihood estimation, it does not require information of the exact distribution of the disturbances. In fact, many common estimators in econometrics can be considered as special cases of GMM.

The theoretical relation that the parameters should satisfy are usually *orthogonality conditions* between some (possibly nonlinear) function of the parameters  $f(\theta)$  and a set of instrumental variables  $z_t$ :

$$E(f(\theta)'Z) = 0, \quad (12.39)$$

where  $\theta$  are the parameters to be estimated. The GMM estimator selects parameter estimates so that the sample correlations between the instruments and the function  $f$  are as close to zero as possible, as defined by the criterion function:

$$J(\theta) = (m(\theta))'Am(\theta), \quad (12.40)$$

where  $m(\theta) = f(\theta)'Z$  and  $A$  is a weighting matrix. Any symmetric positive definite matrix  $A$  will yield a consistent estimate of  $\theta$ . However, it can be shown that a necessary (but not sufficient) condition to obtain an (asymptotically) efficient estimate of  $\theta$  is to set  $A$  equal to the inverse of the covariance matrix of the sample moments  $m$ .

Many standard estimators, including all of the system estimators provided in EViews, can be set up as special cases of GMM. For example, the ordinary least squares estimator can be viewed as a GMM estimator, based upon the conditions that each of the right-hand variables is uncorrelated with the residual.

### Estimation by GMM in EViews

To estimate an equation by GMM, either create a new equation object by selecting **Object/New Object/Equation**, or press the **Estimate** button in the toolbar of an existing equation. From the Equation Specification dialog choose **Estimation Method: GMM**. The estimation specification dialog will change as depicted below.

To obtain GMM estimates in EViews, you need to write the moment condition as an orthogonality condition between an expression including the parameters and a set of instrumental variables. There are two ways you can write the orthogonality condition: with and without a dependent variable.

If you specify the equation either by listing variable names or by an expression with an equals sign, EViews will interpret the moment condition as an orthogonality condition between the instruments and the residuals defined by the equation. If you specify the

equation by an expression without an equals sign, EViews will orthogonalize that expression to the set of instruments.

You must also list the names of the instruments in the **Instrument list** field box of the Equation Specification dialog box. For the GMM estimator to be identified, there must be at least as many instrumental variables as there are parameters to estimate. EViews will always include the constant in the list of instruments.

For example, if you type

Equation Specification:  $y \ c \ x$   
Instrument list:  $c \ z \ w$

the orthogonality conditions given by

$$\begin{aligned}\sum (y_t - c(1) - c(2)x_t) &= 0 \\ \sum (y_t - c(1) - c(2)x_t)z_t &= 0 \\ \sum (y_t - c(1) - c(2)x_t)w_t &= 0\end{aligned}\tag{12.41}$$

If you enter an expression

Equation Specification:  $c(1) * \log(y) + x^{c(2)}$   
Instrument list:  $c \ z \ z(-1)$

the orthogonality conditions are

$$\begin{aligned}\sum (c(1)\log y_t + x_t^{c(2)}) &= 0 \\ \sum (c(1)\log y_t + x_t^{c(2)})z_t &= 0 \\ \sum (c(1)\log y_t + x_t^{c(2)})z_{t-1} &= 0\end{aligned}\tag{12.42}$$

On the right part of the Equation Specification dialog are the options for selecting the weighting matrix  $A$  in the objective function. If you select **Weighting Matrix: Cross section (White Cov)**, the GMM estimates will be robust to heteroskedasticity of unknown form.

If you select **Weighting Matrix: Time series (HAC)**, the GMM estimates will be robust to heteroskedasticity and autocorrelation of unknown form. For the HAC option, you have to specify the a kernel type and bandwidth.

- The **Kernel Options** determine the functional form of the kernel used to weight the autocovariances in computing the weighting matrix.
- The **Bandwidth Selection** option determines how the weights given by the kernel change with the lags of the autocovariances in the computation of the weighting matrix. If you select **Fixed** bandwidth, you may either enter a number for the bandwidth or type `nw` to use Newey and West's fixed bandwidth selection criterion.

- The **Prewhitening** option runs a preliminary VAR(1) prior to estimation to “soak up” the correlation in the moment conditions.

The Technical Notes at the end of Chapter 18 describes these options in more detail.

### Example

Tauchen (1986) considers the problem of estimating the taste parameters  $\beta$ ,  $\gamma$  from the Euler equation

$$(\beta R_{t+1} w_{t+1}^{-\gamma} - 1)' z_t = 0 \tag{12.43}$$

where we use instruments  $z_t = (1, w_t, w_{t-1}, r_t, r_{t-1})'$ . To estimate the parameters  $\beta$ ,  $\gamma$  by GMM, fill in the Equation Specification dialog as

Equation Specification:    `c (1) *r (+1) *w (+1) ^ (-c (2)) -1`  
Instrument list:            `c w w (-1) r r (-1)`

The estimation result using the default HAC **Weighting Matrix** option looks as follows:

Dependent Variable: Implicit Equation  
Method: Generalized Method of Moments  
Date: 09/26/97    Time: 14:02  
Sample(adjusted): 1891 1982  
Included observations: 92 after adjusting endpoints  
No prewhitening  
Bandwidth: Fixed (3)  
Kernel: Bartlett  
Convergence achieved after: 7 weight matrices, 7 total coef iterations  
C(1)\*R(+1)\*W(+1)^(-C(2))-1  
Instrument list: C W W(-1) R R(-1)

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	0.934096	0.018751	49.81600	0.0000
C(2)	1.366396	0.741802	1.841995	0.0688
S.E. of regression	0.154084	Sum squared resid	2.136760	
Durbin-Watson stat	1.903837	J-statistic	0.054523	

Note that when you specify an equation without a dependent variable, EViews does not report some of the regression statistics such as the R-squared. The J-statistic reported at the bottom of the table is the minimized value of the objective function. The J-statistic can be used to carry out hypothesis tests from GMM estimation; see Newey and West (1987a). A simple application of the J-statistic is to test the validity of overidentifying restrictions when you have more instruments than parameters to estimate. In this example, we have five instruments to estimate two parameters and so there are three overidentifying restrictions. Under the null hypothesis that the overidentifying restrictions are satisfied, the J-statistic times the number of regression observations is asymptotically  $\chi^2$  with degrees of freedom equal to the number of overidentifying restrictions. You can compute the test statistic as a named scalar in EViews using the commands



```
scalar overid=eq_gmm.@regobs*eq_gmm.@jstat
scalar overid_p=1-@cchisq(overid,3)
```

where EQ\_GMM is the name of the equation containing the GMM estimates. The second command computes the  $p$ -value of the test statistic as a named scalar OVERID\_P. To view the value of OVERID\_P, double click on its name; the value will be displayed in the status line at the bottom of the EViews window.

## Commands

To estimate an equation by weighted least squares, specify the weighting series in parentheses with the `w=` option after the `ls` command:

```
eq1.ls(w=1/pop) cs c gdp cpi
```

To estimate an equation by two stage least squares, follow the `tsls` command with the dependent variable, the independent variables, an `@` sign, and a list of instruments:

```
equation eq2.tsls cs c gdp @ c cs(-1) gdp(-1)
```

To estimate an equation by GMM, follow the `gmm` command with the dependent variable, the independent variables, an `@` sign, and a list of instruments that define the orthogonality conditions:

```
equation eq3.gmm cs c gdp @ c cs(-1) gdp(-1)
```

You can set the starting values prior to nonlinear estimation by the command

```
param c(1) 1.5 c(2) 1
```

or

```
c(1) = 1.5
c(2) = 1
```

To declare a coefficient vector, specify the number of rows in parentheses and provide a name:

```
coef(4) beta
```

declares a four element coefficient vector named BETA filled with zeros.

See the *Command and Programming Reference* for a complete list of commands and options for single equation estimation in EViews.



## Chapter 13. Time Series Regression

---

In this chapter we discuss single equation regression techniques that are important for the analysis of time series data: testing for serial correlation, estimation of ARMA models, using polynomial distributed lags, and testing for unit root in potentially non-stationary time series.

The chapter focuses on the specification and estimation of time series models. A number of related topics are discussed elsewhere: standard multiple regression techniques are discussed in Chapters 11 and 12, forecasting and inference are discussed extensively in Chapters 14 and 15, vector autoregressions are discussed in Chapter 20, and state space models and the Kalman filter are discussed in Chapter 21.

### Serial Correlation Theory

A common finding in time series regressions is that the residuals are correlated with their own lagged values. This serial correlation violates the standard assumption of regression theory that disturbances are not correlated with other disturbances. The primary problems associated with serial correlation are:

- OLS is no longer efficient among linear estimators. Furthermore, since prior residuals help to predict current residuals, we can take advantage of this information to form a better prediction of the dependent variable.
- Standard errors computed using the textbook OLS formula are not correct, and are generally understated.
- If there are lagged dependent variables on the right-hand side, OLS estimates are biased and inconsistent.

EViews provides tools for detecting serial correlation and estimation methods that take account of its presence.

In general, we will be concerned with specifications of the form:

$$\begin{aligned}y_t &= x_t' \beta + u_t \\ u_t &= z_{t-1}' \gamma + \epsilon_t\end{aligned}\tag{13.1}$$

where  $x_t$  is a vector of explanatory variables observed at time  $t$ ,  $z_{t-1}$  is a vector of variables known in the previous period,  $\beta$  and  $\gamma$  are vectors of parameters,  $u_t$  is a disturbance term, and  $\epsilon_t$  is the innovation in the disturbance. The vector  $z_{t-1}$  may contain lagged values of  $u$ , lagged values of  $\epsilon$ , or both.

The disturbance  $u_t$  is termed the *unconditional residual*. It is the residual based on the structural component ( $x_t'\beta$ ) but not using the information contained in  $z_{t-1}$ . The innovation  $\epsilon_t$  is also known as the *one-period-ahead forecast error* or the *prediction error*. It is the difference between the actual value of the dependent variable and a forecast made on the basis of the independent variables and the past forecast errors.

### The First-Order Autoregressive Model

The simplest and most widely used model of serial correlation is the first-order autoregressive, or AR(1), model. The AR(1) model is specified as

$$\begin{aligned} y_t &= x_t'\beta + u_t \\ u_t &= \rho u_{t-1} + \epsilon_t \end{aligned} \quad (13.2)$$

The parameter  $\rho$  is the first-order serial correlation coefficient. In effect, the AR(1) model incorporates the residual from the past observation into the regression model for the current observation.

### Higher-Order Autoregressive Models

More generally, a regression with an autoregressive process of order  $p$ , AR( $p$ ) error is given by

$$\begin{aligned} y_t &= x_t'\beta + u_t \\ u_t &= \rho_1 u_{t-1} + \rho_2 u_{t-2} + \dots + \rho_p u_{t-p} + \epsilon_t \end{aligned} \quad (13.3)$$

The autocorrelations of a stationary AR( $p$ ) process gradually die out to zero, while the partial autocorrelations for lags larger than  $p$  are zero.

## Testing for Serial Correlation

Before you use an estimated equation for statistical inference (e.g. hypothesis tests and forecasting), you should generally examine the residuals for evidence of serial correlation. EViews provides several methods of testing a specification for the presence of serial correlation.

### The Durbin-Watson Statistic

EViews reports the Durbin-Watson (DW) statistic as a part of the standard regression output. The Durbin-Watson statistic is a test for first-order serial correlation. More formally, the DW statistic measures the linear association between adjacent residuals from a regression model. The Durbin-Watson is a test of the hypothesis  $\rho = 0$  in the specification:

$$u_t = \rho u_{t-1} + \epsilon_t. \quad (13.4)$$

If there is no serial correlation, the DW statistic will be around 2. The DW statistic will fall below 2 if there is positive serial correlation (in the worst case, it will be near zero). If there is negative correlation, the statistic will lie somewhere between 2 and 4.

Positive serial correlation is the most commonly observed form of dependence. As a rule of thumb, with 50 or more observations and only a few independent variables, a DW statistic below about 1.5 is a strong indication of positive first order serial correlation. See Johnston and DiNardo (1997, Chapter 6.6.1) for a thorough discussion on the Durbin-Watson test and a table of the significance points of the statistic.

There are three main limitations of the DW test as a test for serial correlation. First, the distribution of the DW statistic under the null hypothesis depends on the data matrix  $x$ . The usual approach to handling this problem is to place bounds on the critical region, creating a region where the test results are inconclusive. Second, if there are lagged dependent variables on the right-hand side of the regression, the DW test is no longer valid. Lastly, you may only test the null hypothesis of no serial correlation against the alternative hypothesis of first-order serial correlation.

Two other tests of serial correlation—the  $Q$ -statistic and the Breusch-Godfrey LM test—overcome these limitations, and are preferred in most applications.

### Correlograms and $Q$ -statistics

If you select **View/Residual Tests/Correlogram- $Q$ -statistics** on the equation toolbar, EViews will display the autocorrelation and partial autocorrelation functions of the residuals, together with the Ljung-Box  $Q$ -statistics for high-order serial correlation. If there is no serial correlation in the residuals, the autocorrelations and partial autocorrelations at all lags should be nearly zero, and all  $Q$ -statistics should be insignificant with large  $p$ -values.

Note that the  $p$ -values of the  $Q$ -statistics will be computed with the degrees of freedom adjusted for the inclusion of ARMA terms in your regression. There is evidence that some care should be taken in interpreting the results of a Ljung-Box test applied to the residuals from an ARMAX specification (see Dezhbaksh, 1990, for simulation evidence on the finite sample performance of the test in this setting).

Details on the computation of correlograms and  $Q$ -statistics are provided in greater detail in Chapter 7, “Series”, beginning on page 159.

### Serial Correlation LM Test

Selecting **View/Residual Tests/Serial Correlation LM Test...** carries out the Breusch-Godfrey Lagrange multiplier test for general, high-order, ARMA errors. In the Lag Specification dialog box, you should enter the highest order of serial correlation to be tested.

The null hypothesis of the test is that there is no serial correlation in the residuals up to the specified order. EViews reports a statistic labeled “ $F$ -statistic” and an “Obs\*R-squared” ( $NR^2$ —the number of observations times the R-square) statistic. The  $NR^2$  statistic has an asymptotic  $\chi^2$  distribution under the null hypothesis. The distribution of the  $F$ -statistic is not known, but is often used to conduct an informal test of the null.

See Chapter 15, “Specification and Diagnostic Tests”, beginning on page 338 for further discussion of the serial correlation LM test.

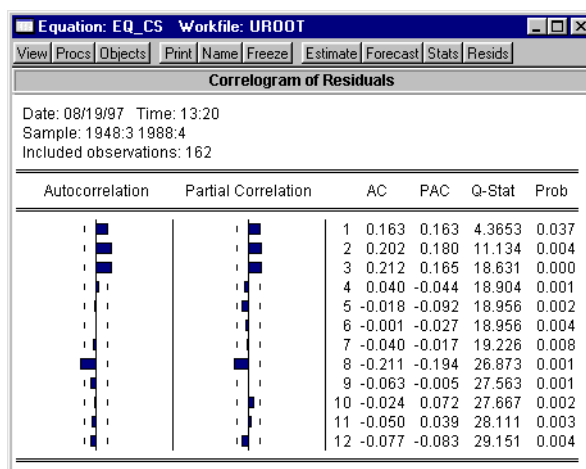
### Example

As an example of the application of these testing procedures, consider the following results from estimating a simple consumption function by ordinary least squares:

Dependent Variable: CS				
Method: Least Squares				
Date: 08/19/97 Time: 13:03				
Sample: 1948:3 1988:4				
Included observations: 162				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-9.227624	5.898177	-1.564487	0.1197
GDP	0.038732	0.017205	2.251193	0.0257
CS(-1)	0.952049	0.024484	38.88516	0.0000
R-squared	0.999625	Mean dependent var	1781.675	
Adjusted R-squared	0.999621	S.D. dependent var	694.5419	
S.E. of regression	13.53003	Akaike info criterion	8.066045	
Sum squared resid	29106.82	Schwarz criterion	8.123223	
Log likelihood	-650.3497	F-statistic	212047.1	
Durbin-Watson stat	1.672255	Prob(F-statistic)	0.000000	

A quick glance at the results reveals that the coefficients are statistically significant and the fit is very tight. However, if the error term is serially correlated, the estimated OLS standard errors are invalid and the estimated coefficients will be biased and inconsistent due to the presence of a lagged dependent variable on the right-hand side. The Durbin-Watson statistic is not appropriate as a test for serial correlation in this case, since there is a lagged dependent variable on the right-hand side of the equation.

Selecting **View/Residual Tests/Correlogram-Q-statistics** from this equation produces the following view:



The correlogram has spikes at lags up to three and at lag eight. The  $Q$ -statistics are significant at all lags, indicating significant serial correlation in the residuals.

Selecting **View/Residual Tests/Serial Correlation LM Test...** and entering a lag of 4 yields the following result:

Breusch-Godfrey Serial Correlation LM Test:			
F-statistic	3.654696	Probability	0.007109
Obs*R-squared	13.96215	Probability	0.007417

The test rejects the hypothesis of no serial correlation up to order four. The  $Q$ -statistic and the LM test both indicate that the residuals are serially correlated and the equation should be re-specified before using it for hypothesis tests and forecasting.

## Estimating AR Models

Before you use the tools described in this section, you may first wish to examine your model for other signs of misspecification. Serial correlation in the errors may be evidence of serious problems with your specification. In particular, you should be on guard for an excessively restrictive specification that you arrived at by experimenting with ordinary least squares. Sometimes, adding improperly excluded variables to your regression will eliminate the serial correlation.

For a discussion of the efficiency gains from the serial correlation correction and some Monte-Carlo evidence, see Rao and Griliches (1969).

### First-Order Serial Correlation

To estimate an AR(1) model in EViews, open an equation by selecting **Quick/Estimate Equation...** and enter your specification as usual, adding the expression “AR(1)” to the end of your list. For example, to estimate a simple consumption function with AR(1) errors,

$$\begin{aligned}CS_t &= c_1 + c_2 GDP_t + u_t \\ u_t &= \rho u_{t-1} + \epsilon_t\end{aligned}\tag{13.5}$$

you should specify your equation as

```
cs c gdp ar(1)
```

EViews automatically adjusts your sample to account for the lagged data used in estimation, estimates the model, and reports the adjusted sample along with the remainder of the estimation output.

### Higher-Order Serial Correlation

Estimating higher order AR models is only slightly more complicated. To estimate an AR( $k$ ), you should enter your specification, followed by expressions for each AR term you wish to include. If you wish to estimate a model with autocorrelations from one to five:

$$\begin{aligned}CS_t &= c_1 + c_2 GDP_t + u_t \\ u_t &= \rho_1 u_{t-1} + \rho_2 u_{t-2} + \dots + \rho_5 u_{t-5} + \epsilon_t\end{aligned}\tag{13.6}$$

you should enter

```
cs c gdp ar(1) ar(2) ar(3) ar(4) ar(5)
```

By requiring that you enter all of the autocorrelations you wish to include in your model, EViews allows you great flexibility in restricting lower order correlations to be zero. For example, if you have quarterly data and want to include a single term to account for seasonal autocorrelation, you could enter

```
cs c gdp ar(4)
```

### Nonlinear Models with Serial Correlation

EViews can estimate nonlinear regression models with additive AR errors. For example, suppose you wish to estimate the following nonlinear specification with an AR(2) error:

$$\begin{aligned}CS_t &= c_1 + GDP_t^{c_2} + u_t \\ u_t &= c_3 u_{t-1} + c_4 u_{t-2} + \epsilon_t\end{aligned}\tag{13.7}$$



Simply specify your model using EViews expressions, followed by an additive term describing the AR correction enclosed in square brackets. The AR term should contain a coefficient assignment for each AR lag, separated by commas:

$$cs = c(1) + gdp \wedge c(2) + [ar(1)=c(3), ar(2)=c(4)]$$

EViews transforms this nonlinear model by differencing, and estimates the transformed nonlinear specification using a Gauss-Newton iterative procedure (see “How EViews Estimates AR Models” on page 280, below).

## Two-Stage Regression Models with Serial Correlation

By combining two-stage least squares or two-stage nonlinear least squares with AR terms, you can estimate models where there is correlation between regressors and the innovations as well as serial correlation in the residuals.

If the original regression model is linear, EViews uses the Marquardt algorithm to estimate the parameters of the transformed specification. If the original model is nonlinear, EViews uses Gauss-Newton to estimate the AR corrected specification.

For further details on the algorithms and related issues associated with the choice of instruments, see the discussion in “TSLS with AR errors”, beginning on page 256.

## Output from AR Estimation

When estimating an AR model, some care must be taken in interpreting your results. While the estimated coefficients, coefficient standard errors, and  $t$ -statistics may be interpreted in the usual manner, results involving residuals differ from those computed in OLS settings.

To understand these differences, keep in mind that there are two different residuals associated with an AR model. The first are the estimated *unconditional residuals*,

$$\hat{u}_t = y_t - x_t' b, \quad (13.8)$$

which are computed using the original variables, and the estimated coefficients,  $b$ . These residuals are the errors that you would observe if you made a prediction of the value of  $y_t$  using contemporaneous information, but ignoring the information contained in the lagged residual.

Normally, there is no strong reason to examine these residuals, and EViews does not automatically compute them following estimation.

The second set of residuals are the estimated *one-period-ahead forecast errors*,  $\hat{\epsilon}$ . As the name suggests, these residuals represent the forecast errors you would make if you computed forecasts using a prediction of the residuals based upon past values of your data, in addition to the contemporaneous information. In essence, you improve upon the unconditional forecasts and residuals by taking advantage of the predictive power of the lagged residuals.

For AR models, the residual-based regression statistics—such as the  $R^2$ , the standard error of regression, and the Durbin-Watson statistic—reported by EViews are based on the one-period-ahead forecast errors,  $\hat{\epsilon}$ .

A set of statistics that is unique to AR models is the estimated AR parameters,  $\hat{\rho}_i$ . For the simple AR(1) model, the estimated parameter  $\hat{\rho}$  is the serial correlation coefficient of the unconditional residuals. For a stationary AR(1) model, the true  $\rho$  lies between  $-1$  (extreme negative serial correlation) and  $+1$  (extreme positive serial correlation). The stationarity condition for general AR( $p$ ) processes is that the inverted roots of the lag polynomial lie inside the unit circle. EViews reports these roots as **Inverted AR Roots** at the bottom of the regression output. There is no particular problem if the roots are imaginary, but a stationary AR model should have all roots with modulus less than one.

### How EViews Estimates AR Models

Textbooks often describe techniques for estimating AR models. The most widely discussed approaches, the Cochrane-Orcutt, Prais-Winsten, Hatanaka, and Hildreth-Lu procedures, are multi-step approaches designed so that estimation can be performed using standard *linear* regression. All of these approaches suffer from important drawbacks which occur when working with models containing lagged dependent variables as regressors, or models using higher-order AR specifications; see Davidson and MacKinnon, (1994, pp. 329–341), Greene (1997, p. 600–607).

EViews estimates AR models using nonlinear regression techniques. This approach has the advantage of being easy to understand, generally applicable, and easily extended to nonlinear specifications and models that contain endogenous right-hand side variables. Note that the nonlinear least squares estimates are asymptotically equivalent to maximum likelihood estimates and are asymptotically efficient.

To estimate an AR(1) model, EViews transforms the linear model

$$\begin{aligned} y_t &= x_t' \beta + u_t \\ u_t &= \rho u_{t-1} + \epsilon_t \end{aligned} \quad (13.9)$$

into the nonlinear model,

$$y_t = \rho y_{t-1} + (x_t - \rho x_{t-1})' \beta + \epsilon_t, \quad (13.10)$$

by substituting the second equation into the first, and rearranging terms. The coefficients  $\rho$  and  $\beta$  are estimated simultaneously by applying a Marquardt nonlinear least squares algorithm to the transformed equation. See Appendix E for details on nonlinear estimation.

For a nonlinear AR(1) specification, EViews transforms the nonlinear model

$$\begin{aligned} y_t &= f(x_t, \beta) + u_t \\ u_t &= \rho u_{t-1} + \epsilon_t \end{aligned} \quad (13.11)$$

into the alternative nonlinear specification

$$y_t = \rho y_{t-1} + f(x_t, \beta) - \rho f(x_{t-1}, \beta) + \epsilon_t \quad (13.12)$$

and estimates the coefficients using a Marquardt nonlinear least squares algorithm.

Higher order AR specifications are handled analogously. For example, a nonlinear AR(3) is estimated using nonlinear least squares on the equation

$$y_t = (\rho_1 y_{t-1} + \rho_2 y_{t-2} + \rho_3 y_{t-3}) + f(x_t, \beta) - \rho_1 f(x_{t-1}, \beta) - \rho_2 f(x_{t-2}, \beta) - \rho_3 f(x_{t-3}, \beta) + \epsilon_t \quad (13.13)$$

For details, see Fair (1984, pp. 210–214), and Davidson and MacKinnon (1996, pp. 331–341).

## ARIMA Theory

ARIMA (autoregressive integrated moving average) models are generalizations of the simple AR model that use three tools for modeling the serial correlation in the disturbance:

- The first tool is the autoregressive, or AR, term. The AR(1) model introduced above uses only the first-order term but, in general, you may use additional, higher-order AR terms. Each AR term corresponds to the use of a lagged value of the residual in the forecasting equation for the unconditional residual. An autoregressive model of order  $p$ , AR( $p$ ) has the form

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \dots + \rho_p u_{t-p} + \epsilon_t. \quad (13.14)$$

- The second tool is the integration order term. Each integration order corresponds to differencing the series being forecast. A first-order integrated component means that the forecasting model is designed for the first difference of the original series. A second-order component corresponds to using second differences, and so on.
- The third tool is the MA, or moving average term. A moving average forecasting model uses lagged values of the forecast error to improve the current forecast. A first-order moving average term uses the most recent forecast error, a second-order term uses the forecast error from the two most recent periods, and so on. An MA( $q$ ) has the form:

$$u_t = \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \dots + \theta_q \epsilon_{t-q}. \quad (13.15)$$

Please be aware that some authors and software packages use the opposite sign convention for the  $\theta$  coefficients so that the signs of the MA coefficients may be reversed.

The autoregressive and moving average specifications can be combined to form an ARMA( $p, q$ ) specification

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \dots + \rho_p u_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \dots + \theta_q \epsilon_{t-q} \quad (13.16)$$

Although econometricians typically use ARIMA models applied to the residuals from a regression model, the specification can also be applied directly to a series. This latter approach provides a univariate model, specifying the conditional mean of the series as a constant, and measuring the residuals as differences of the series from its mean.

### Principles of ARIMA Modeling (Box-Jenkins 1976)

In ARIMA forecasting, you assemble a complete forecasting model by using combinations of the three building blocks described above. The first step in forming an ARIMA model for a series of residuals is to look at its autocorrelation properties. You can use the correlogram view of a series for this purpose, as outlined in Chapter 7, “Correlograms and Q-statistics” on page 275.

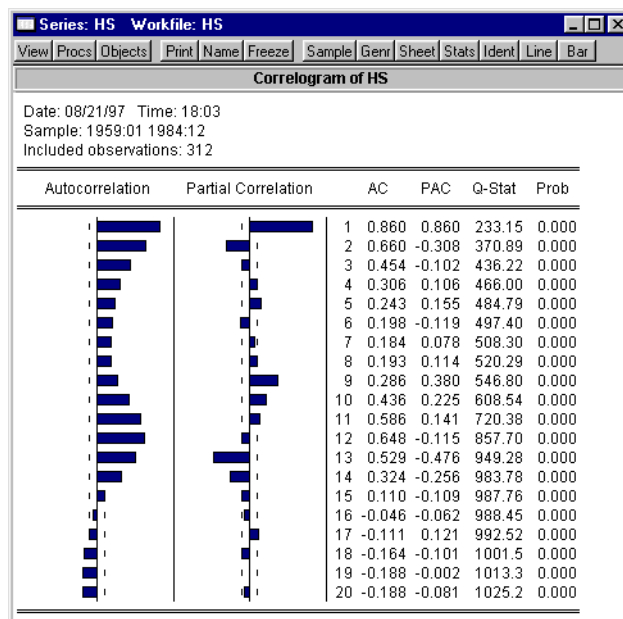
This phase of the ARIMA modeling procedure is called *identification* (not to be confused with the same term used in the simultaneous equations literature). The nature of the correlation between current values of residuals and their past values provides guidance in selecting an ARIMA specification.

The autocorrelations are easy to interpret—each one is the correlation coefficient of the current value of the series with the series lagged a certain number of periods. The partial autocorrelations are a bit more complicated; they measure the correlation of the current and lagged series after taking into account the predictive power of all the values of the series with smaller lags. The partial autocorrelation for lag 6, for example, measures the added predictive power of  $u_{t-6}$  when  $u_1, \dots, u_{t-5}$  are already in the prediction model. In fact, the partial autocorrelation is precisely the regression coefficient of  $u_{t-6}$  in a regression where the earlier lags are also used as predictors of  $u_t$ .

If you suspect that there is a distributed lag relationship between your dependent (left-hand) variable and some other predictor, you may want to look at their cross correlations before carrying out estimation.

The next step is to decide what kind of ARIMA model to use. If the autocorrelation function dies off smoothly at a geometric rate, and the partial autocorrelations were zero after one lag, then a first-order autoregressive model is appropriate. Alternatively, if the autocorrelations were zero after one lag and the partial autocorrelations declined geometrically, a first-order moving average process would seem appropriate. If the autocorrelations appear to have a seasonal pattern, this would suggest the presence of a seasonal ARMA structure (see “Seasonal ARMA Terms” on page 286 below).

For example, we can examine the correlogram of the DRI Basics housing series in the HS.WF1 workfile by selecting **View/Correlogram...** from the HS series toolbar:



The “wavy” cyclical correlogram with a seasonal frequency suggests fitting a seasonal ARMA model to HS.

The goal of ARIMA analysis is a parsimonious representation of the process governing the residual. You should use only enough AR and MA terms to fit the properties of the residuals. The Akaike information criterion and Schwarz criterion provided with each set of estimates may also be used as a guide for the appropriate lag order selection.

After fitting a candidate ARIMA specification, you should verify that there are no remaining autocorrelations that your model has not accounted for. Examine the autocorrelations and the partial autocorrelations of the innovations (the residuals from the ARIMA model) to see if any important forecasting power has been overlooked. EViews provides views for diagnostic checks after estimation.

## Estimating ARIMA Models

EViews estimates general ARIMA specifications that allow for right-hand side explanatory variables. Despite the fact that these models are sometimes termed ARIMAX specifications, we will refer to this general class of models as ARIMA.

To specify your ARIMA model, you will:

- Difference your dependent variable, if necessary, to account for the order of integration.

- Describe your structural regression model (dependent variables and regressors) and add any AR or MA terms, as described above.

### Differencing

The `d` operator can be used to specify differences of series. To specify first differencing, simply include the series name in parentheses after `d`. For example, `d(gdp)` specifies the first difference of GDP, or `GDP-GDP(-1)`.

More complicated forms of differencing may be specified with two optional parameters,  $n$  and  $s$ . `d(x, n)` specifies the  $n$ -th order difference of the series  $X$ :

$$d(x, n) = (1 - L)^n x, \quad (13.17)$$

where  $L$  is the lag operator. For example, `d(gdp, 2)` specifies the second order difference of GDP:

$$d(gdp, 2) = gdp - 2 * gdp(-1) + gdp(-2)$$

`d(x, n, s)` specifies  $n$ -th order ordinary differencing of  $X$  with a seasonal difference at lag  $s$ :

$$d(x, n, s) = (1 - L)^n (1 - L^s) x. \quad (13.18)$$

For example, `d(gdp, 0, 4)` specifies zero ordinary differencing with a seasonal difference at lag 4, or `GDP-GDP(-4)`.

If you need to work in logs, you can also use the `dlog` operator, which returns differences in the log values. For example, `dlog(gdp)` specifies the first difference of  $\log(\text{GDP})$  or  $\log(\text{GDP}) - \log(\text{GDP}(-1))$ . You may also specify the  $n$  and  $s$  options as described for the simple `d` operator, `dlog(x, n, s)`.

There are two ways to estimate integrated models in EViews. First, you may generate a new series containing the differenced data, and then estimate an ARMA model using the new data. For example, to estimate a Box-Jenkins ARIMA(1, 1, 1) model for M1, you can enter:

```
series dm1 = d(m1)
ls dm1 c ar(1) ma(1)
```

Alternatively, you may include the difference operator `d` directly in the estimation specification. For example, the same ARIMA(1,1,1) model can be estimated by the one-line command

```
ls d(m1) c ar(1) ma(1)
```

The latter method should generally be preferred for an important reason. If you define a new variable, such as DM1 above, and use it in your estimation procedure, then when you forecast from the estimated model, EViews will make forecasts of the dependent variable DM1. That is, you will get a forecast of the differenced series. If you are really interested in forecasts of the level variable, in this case M1, you will have to manually transform the forecasted value and

adjust the computed standard errors accordingly. Moreover, if any other transformation or lags of M1 are included as regressors, EViews will not know that they are related to DM1. If, however, you specify the model using the difference operator expression for the dependent variable,  $d(m1)$ , the forecasting procedure will provide you with the option of forecasting the level variable, in this case M1.

The difference operator may also be used in specifying exogenous variables and can be used in equations without ARMA terms. Simply include them in the list of regressors in addition to the endogenous variables. For example,

```
d(cs,2) c d(gdp,2) d(gdp(-1),2) d(gdp(-2),2) time
```

is a valid specification that employs the difference operator on both the left-hand and right-hand sides of the equation.

## ARMA Terms

The AR and MA parts of your model will be specified using the keywords `ar` and `ma` as part of the equation. We have already seen examples of this approach in our specification of the AR terms above, and the concepts carry over directly to MA terms.

For example, to estimate a second-order autoregressive and first-order moving average error process ARMA(2,1), you would include expressions for the AR(1), AR(2), and MA(1) terms along with your other regressors:

```
c gov ar(1) ar(2) ma(1)
```

Once again, you need not use the AR and MA terms consecutively. For example, if you want to fit a fourth-order autoregressive model to take account of seasonal movements, you could use AR(4) by itself:

```
c gov ar(4)
```

You may also specify a pure moving average model by using only MA terms. Thus,

```
c gov ma(1) ma(2)
```

indicates an MA(2) model for the residuals.

The traditional Box-Jenkins or ARMA models do not have any right-hand side variables except for the constant. In this case, your list of regressors would just contain a C in addition to the AR and MA terms. For example,

```
c ar(1) ar(2) ma(1) ma(2)
```

is a standard Box-Jenkins ARMA (2,2).

## Seasonal ARMA Terms

Box and Jenkins (1976) recommend the use of seasonal autoregressive (SAR) and seasonal moving average (SMA) terms for monthly or quarterly data with systematic seasonal movements. A SAR( $p$ ) term can be included in your equation specification for a seasonal autoregressive term with lag  $p$ . The lag polynomial used in estimation is the product of the one specified by the AR terms and the one specified by the SAR terms. The purpose of the SAR is to allow you to form the product of lag polynomials.

Similarly, SMA( $q$ ) can be included in your specification to specify a seasonal moving average term with lag  $q$ . The lag polynomial used in estimation is the product of the one defined by the MA terms and the one specified by the SMA terms. As with the SAR, the SMA term allows you to build up a polynomial that is the product of underlying lag polynomials.

For example, a second-order AR process without seasonality is given by

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \epsilon_t, \quad (13.19)$$

which can be represented using the lag operator  $L$ ,  $L^n x_t = x_{t-n}$  as

$$(1 - \rho_1 L - \rho_2 L^2)u_t = \epsilon_t. \quad (13.20)$$

You can estimate this process by including `ar(1)` and `ar(2)` terms in the list of regressors. With quarterly data, you might want to add a `sar(4)` expression to take account of seasonality. If you specify the equation as

```
sales c inc ar(1) ar(2) sar(4)
```

then the estimated error structure would be:

$$(1 - \rho_1 L - \rho_2 L^2)(1 - \phi L^4)u_t = \epsilon_t. \quad (13.21)$$

The error process is equivalent to:

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \phi u_{t-4} - \phi \rho_1 u_{t-5} - \phi \rho_2 u_{t-6} + \epsilon_t. \quad (13.22)$$

The parameter  $\phi$  is associated with the seasonal part of the process. Note that this is an AR(6) process with nonlinear restrictions on the coefficients.

As another example, a second-order MA process without seasonality may be written

$$u_t = \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2}, \quad (13.23)$$

or using lag operators,

$$u_t = (1 + \theta_1 L + \theta_2 L^2)\epsilon_t. \quad (13.24)$$

You can estimate this second-order process by including both the MA(1) and MA(2) terms in your equation specification.



With quarterly data, you might want to add `sma(4)` to take account of seasonality. If you specify the equation as

```
cs c ad ma(1) ma(2) sma(4)
```

then the estimated model is:

$$\begin{aligned} CS_t &= \beta_1 + \beta_2 AD_t + u_t \\ u_t &= (1 + \theta_1 L + \theta_2 L^2)(1 + \omega L^4)\epsilon_t \end{aligned} \quad (13.25)$$

The error process is equivalent to

$$u_t = \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \omega \epsilon_{t-4} + \omega \theta_1 \epsilon_{t-5} + \omega \theta_2 \epsilon_{t-6}. \quad (13.26)$$

The parameter  $\omega$  is associated with the seasonal part of the process. This is just an MA(6) process with nonlinear restrictions on the coefficients. You can also include both SAR and SMA terms.

## Output from ARIMA Estimation

The output from estimation with AR or MA specifications is the same as for ordinary least squares, with the addition of a lower block that shows the reciprocal roots of the AR and MA polynomials. If we write the general ARMA model using the lag polynomial  $\rho(L)$  and  $\theta(L)$  as

$$\rho(L)u_t = \theta(L)\epsilon_t, \quad (13.27)$$

then the reported roots are the roots of the polynomials

$$\rho(x^{-1}) = 0 \quad \text{and} \quad \theta(x^{-1}) = 0. \quad (13.28)$$

The roots, which may be imaginary, should have modulus no greater than one. The output will display a warning message if any of the roots violate this condition.

If  $\rho$  has a real root whose absolute value exceeds one or a pair of complex reciprocal roots outside the unit circle (that is, with modulus greater than one), it means that the autoregressive process is explosive.

If  $\theta$  has reciprocal roots outside the unit circle, we say that the MA process is *noninvertible*, which makes interpreting and using the MA results difficult. However, noninvertibility poses no substantive problem, since as Hamilton (1994, p. 65) notes, there is always an equivalent representation for the MA model where the reciprocal roots lie inside the unit circle. Accordingly, you should re-estimate your model with different starting values until you get a moving average process that satisfies invertibility. Alternatively, you may wish to turn off MA backcasting (see “Backcasting MA terms” on page 290).

If the estimated MA process has roots with modulus close to one, it is a sign that you may have over-differenced the data. The process will be difficult to estimate and even more difficult to forecast. If possible, you should re-estimate with one less round of differencing.

Consider the following example output from ARMA estimation:

Dependent Variable: R				
Method: Least Squares				
Date: 08/14/97 Time: 16:53				
Sample(adjusted): 1954:06 1993:07				
Included observations: 470 after adjusting endpoints				
Convergence achieved after 25 iterations				
Backcast: 1954:01 1954:05				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	8.614804	0.961559	8.959208	0.0000
AR(1)	0.983011	0.009127	107.7077	0.0000
SAR(4)	0.941898	0.018788	50.13275	0.0000
MA(1)	0.513572	0.040236	12.76402	0.0000
SMA(4)	-0.960399	0.000601	-1598.423	0.0000
R-squared	0.991557	Mean dependent var	6.978830	
Adjusted R-squared	0.991484	S.D. dependent var	2.919607	
S.E. of regression	0.269420	Akaike info criterion	0.225489	
Sum squared resid	33.75296	Schwarz criterion	0.269667	
Log likelihood	-47.98992	F-statistic	13652.76	
Durbin-Watson stat	2.099958	Prob(F-statistic)	0.000000	
Inverted AR Roots	.99	.98		
Inverted MA Roots	.99	-.00+.99i	-.00 -.99i	-.51
	-.99			

This estimation result corresponds to the following specification:

$$y_t = 8.61 + u_t \quad (13.29)$$

$$(1 - 0.87L)(1 - 0.94L^4)u_t = (1 + 0.51L)(1 - 0.96L^4)\epsilon_t$$

or equivalently, to

$$y_t = 0.0012 + 0.98y_{t-1} + 0.94y_{t-4} - 0.92y_{t-5} + \epsilon_t + 0.51\epsilon_{t-1} - 0.96\epsilon_{t-4} - 0.49\epsilon_{t-5} \quad (13.30)$$

Note that the signs of the MA terms may be reversed from those in textbooks. Note also that the inverted roots have moduli very close to one, which is typical for many macro time series models.

## Estimation Options

ARMA estimation employs the same nonlinear estimation techniques described earlier for AR estimation. These nonlinear estimation techniques are discussed further in Chapter 12, “Additional Regression Methods”, beginning on page 260.

You may need to use the **Estimation Options** dialog box to control the iterative process. EViews provides a number of options that allow you to control the iterative procedure of the estimation

algorithm. In general, you can rely on the EViews choices but on occasion you may wish to override the default settings.

### Iteration Limits and Convergence Criterion

Controlling the maximum number of iterations and convergence criterion are described in detail in Appendix E, “Iteration and Convergence options” on page 596.

### Starting Values for ARMA Estimation

As discussed above, models with AR or MA terms are estimated by nonlinear least squares. Nonlinear estimation techniques require starting values for all coefficient estimates. Normally, EViews determines its own starting values and for the most part this is an issue that you need not be concerned about. However, there are a few times when you may want to override the default starting values.

First, estimation will sometimes halt when the maximum number of iterations is reached, despite the fact that convergence is not achieved. Resuming the estimation with starting values from the previous step causes estimation to pick up where it left off instead of starting over. You may also want to try different starting values to ensure that the estimates are a global rather than a local minimum of the squared errors. You might also want to supply starting values if you have a good idea of what the answers should be, and want to speed up the estimation process.

To control the starting values for ARMA estimation, click on the **Options** button in the Equation Specification dialog. Among the options which EViews provides are several alternatives for setting starting values that you can see by accessing the drop-down menu labeled **Starting Coefficient Values for ARMA**.

EViews’ default approach is **OLS/TSLs**, which runs a preliminary estimation without the ARMA terms and then starts nonlinear estimation from those values. An alternative is to use fractions of the OLS or TSLs coefficients as starting values. You can choose **.8**, **.5**, **.3**, or you can start with all coefficient values set equal to zero.

The final starting value option is **User Supplied**. Under this option, EViews uses the coefficient values that are in the coefficient vector. To set the starting values, open a window for the coefficient vector C by double clicking on the icon, and editing the values.

To properly set starting values, you will need a little more information about how EViews assigns coefficients for the ARMA terms. As with other estimation methods, when you specify your equation as a list of variables, EViews uses the built-in C coefficient vector. It assigns coefficient numbers to the variables in the following order:

- First are the coefficients of the variables, in order of entry.
- Next come the AR terms in the order you typed them.

- The SAR, MA, and SMA coefficients follow, in that order.

Thus the following two specifications will have their coefficients in the same order:

```
y c x ma(2) ma(1) sma(4) ar(1)
y sma(4) c ar(1) ma(2) x ma(1)
```

You may also assign values in the C vector using the `param` command:

```
param c(1) 50 c(2) .8 c(3) .2 c(4) .6 c(5) .1 c(6) .5
```

The starting values will be 50 for the constant, .8 for X, .2 for AR(1), .6 for MA(2), .1 for MA(1) and .5 for SMA(4). Following estimation, you can always see the assignment of coefficients by looking at the **Representations** view of your equation.

You can also fill the C vector from any estimated equation (without typing the numbers) by choosing **Procs/Update Coefs from Equation** in the equation toolbar.

### Backcasting MA terms

By default, EViews backcasts MA terms (Box and Jenkins, 1976). Consider an MA( $q$ ) model of the form

$$\begin{aligned} y_t &= X_t' \beta + u_t \\ u_t &= \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \dots + \theta_q \epsilon_{t-q} \end{aligned} \quad (13.31)$$

Given initial values,  $\hat{\beta}$  and  $\hat{\phi}$ , EViews first computes the unconditional residuals  $\hat{u}_t$  for  $t = 1, \dots, T$ , and uses the backward recursion:

$$\tilde{\epsilon}_t = \hat{u}_t - \hat{\phi}_1 \tilde{\epsilon}_{t+1} - \dots - \hat{\phi}_q \tilde{\epsilon}_{t+q} \quad (13.32)$$

to compute backcast values of  $\epsilon$  to  $\epsilon_{-(q-1)}$ . To start this recursion, the  $q$  values for the innovations *beyond* the estimation sample are set to zero:

$$\tilde{\epsilon}_{T+1} = \tilde{\epsilon}_{T+2} = \dots = \tilde{\epsilon}_{T+q} = 0. \quad (13.33)$$

Next, a forward recursion is used to estimate the values of the innovations

$$\hat{\epsilon}_t = \hat{u}_t - \hat{\phi}_1 \hat{\epsilon}_{t-1} - \dots - \hat{\phi}_q \hat{\epsilon}_{t-q}, \quad (13.34)$$

using the backcasted values of the innovations (to initialize the recursion) and the actual residuals. If your model also includes AR terms, EViews will  $\rho$ -difference the  $\hat{u}_t$  to eliminate the serial correlation prior to performing the backcast.

Lastly, the sum of squared residuals (SSR) is formed as a function of the  $\beta$  and  $\phi$ , using the fitted values of the lagged innovations:

$$SSR(\beta, \phi) = \sum_{t=p+1}^T (y_t - X_t' \beta - \phi_1 \hat{\epsilon}_{t-1} - \dots - \phi_q \hat{\epsilon}_{t-q})^2. \quad (13.35)$$

This expression is minimized with respect to  $\beta$  and  $\phi$ .

The backcast step, forward recursion, and minimization procedures, are repeated until the estimates of  $\beta$  and  $\phi$  converge.

If backcasting is turned off, the estimated values of the pre-sample  $\epsilon$  are set to zero:

$$\epsilon_{-(q-1)} = \dots = \hat{\epsilon}_0 = 0, \quad (13.36)$$

and forward recursion is used to solve for the remaining values of the innovations.

### Dealing with Estimation Problems

Since EViews uses nonlinear least squares algorithms to estimate ARMA models, all of the discussion in Chapter 12, “Solving Estimation Problems” on page 266, is applicable, especially the advice to try alternative starting values.

There are a few other issues to consider that are specific to estimation of ARMA models.

First, MA models are notoriously difficult to estimate. In particular, you should avoid high order MA terms unless absolutely required for your model as they are likely to cause estimation difficulties. For example, a single large spike at lag 57 in the correlogram does not necessarily require you to include an MA(57) term in your model unless you know there is something special happening every 57 periods. It is more likely that the spike in the correlogram is simply the product of one or more outliers in the series. By including many MA terms in your model, you lose degrees of freedom, and may sacrifice stability and reliability of your estimates.

If the underlying roots of the MA process have modulus close to one, you may encounter estimation difficulties, with EViews reporting that it cannot improve the sum-of-squares or that it failed to converge in the maximum number of iterations. This behavior may be a sign that you have over-differenced the data. You should check the correlogram of the series to determine whether you can re-estimate with one less round of differencing.

Lastly, if you continue to have problems, you may wish to turn off MA backcasting.

### TSLS with ARIMA errors

Two-stage least squares or instrumental variable estimation with ARIMA poses no particular difficulties.

For a detailed discussion of how to estimate TSLS specifications with ARMA errors, see “Two-Stage Least Squares” on page 253.

### Nonlinear Models with ARMA errors

EViews will estimate nonlinear ordinary and two-stage least squares models with autoregressive error terms. For details, see the extended discussion in “Nonlinear Least Squares” on page 260.

EViews does not currently estimate nonlinear models with MA errors. You can, however, use the state space object to specify and estimate these models (see Chapter 21).

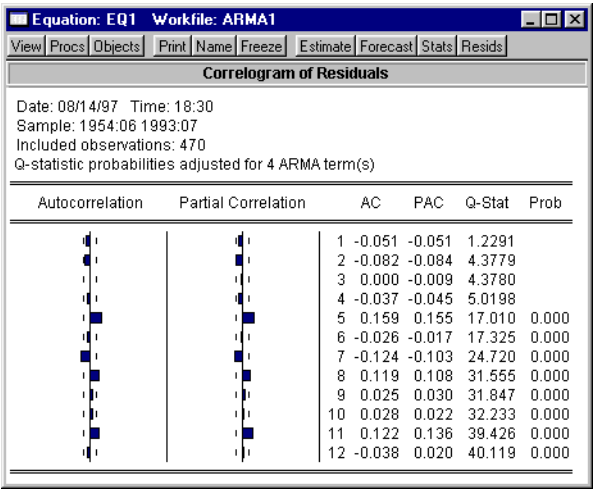
Weighted Models with ARMA errors

EViews does not have procedures to automatically estimate weighted models with ARMA error terms—if you add AR terms to a weighted model, the weighting series will be ignored. You can, of course, always construct the weighted series and then perform estimation using the weighted data and ARMA terms.

Diagnostic Evaluation

If your ARMA model is correctly specified, the residuals from the model should be nearly white noise. This means that there should be no serial correlation left in the residuals. The Durbin-Watson statistic reported in the regression output is a test for AR(1) in the absence of lagged dependent variables on the right-hand side. As discussed above, more general tests for serial correlation in the residuals can be carried out with **View/Residual Tests/Correlogram-Q-statistic** and **View/Residual Tests/Serial Correlation LM Test....**

For the example seasonal ARMA model, the residual correlogram looks as follows:



The correlogram has a significant spike at lag 5 and all subsequent *Q*-statistics are highly significant. This result clearly indicates the need for respecification of the model.

Polynomial Distributed Lags (PDLs)

A distributed lag is a relation of the type

$$y_t = w_t\delta + \beta_0x_t + \beta_1x_{t-1} + \dots + \beta_kx_{t-k} + \epsilon_t \quad (13.37)$$

The coefficients  $\beta$  describe the lag in the effect of  $x$  on  $y$ . In many cases, the coefficients can be estimated directly using this specification. In other cases, the high collinearity of current and lagged values of  $x$  will defeat direct estimation.

You can reduce the number of parameters to be estimated by using polynomial distributed lags (PDLs) to impose a smoothness condition on the lag coefficients. Smoothness is expressed as requiring that the coefficients lie on a polynomial of relatively low degree. A polynomial distributed lag model with order  $p$  restricts the  $\beta$  coefficients to lie on a  $p$ -th order polynomial of the form

$$\beta_j = \gamma_1 + \gamma_2(j - \bar{c}) + \gamma_3(j - \bar{c})^2 + \dots + \gamma_{p+1}(j - \bar{c})^p \quad (13.38)$$

for  $j = 1, \dots, k$ , where  $\bar{c}$  is a pre-specified constant given by

$$\bar{c} = \begin{cases} k/2 & \text{if } p \text{ is even} \\ (k-1)/2 & \text{if } p \text{ is odd} \end{cases} \quad (13.39)$$

The PDL is sometimes referred to as an Almon lag. The constant  $\bar{c}$  is included only to avoid numerical problems that can arise from collinearity and does not affect the estimates of  $\beta$ .

This specification allows you to estimate a model with  $k$  lags of  $x$  using only  $p$  parameters (if you choose  $p > k$ , EViews will return a “Near Singular Matrix” error).

If you specify a PDL, EViews substitutes Equation (13.38) into Equation (13.37) yielding

$$y_t = \alpha + \gamma_1z_1 + \gamma_2z_2 + \dots + \gamma_{p+1}z_{p+1} + \epsilon_t \quad (13.40)$$

where

$$\begin{aligned} z_1 &= x_t + x_{t-1} + \dots + x_{t-k} \\ z_2 &= -\bar{c}x_t + (1 - \bar{c})x_{t-1} + \dots + (k - \bar{c})x_{t-k} \\ &\vdots \\ z_{p+1} &= (-\bar{c})^p x_t + (1 - \bar{c})^p x_{t-1} + \dots + (k - \bar{c})^p x_{t-k} \end{aligned} \quad (13.41)$$

Once we estimate  $\gamma$  from Equation (13.40), we can recover the parameters of interest  $\beta$ , and their standard errors using the relationship described in Equation (13.38). This procedure is straightforward since  $\beta$  is a linear transformation of  $\gamma$ .

The specification of a polynomial distributed lag has three elements: the length of the lag  $k$ , the degree of the polynomial (the highest power in the polynomial)  $p$ , and the constraints that you want to apply. A near end constraint restricts the one-period lead effect of  $x$  on  $y$  to be zero:

$$\beta_{-1} = \gamma_1 + \gamma_2(-1 - \bar{c}) + \dots + \gamma_{p+1}(-1 - \bar{c})^p = 0. \quad (13.42)$$

A far end constraint restricts the effect of  $x$  on  $y$  to die off beyond the number of specified lags:

$$\beta_{k+1} = \gamma_1 + \gamma_2(k+1-\bar{c}) + \dots + \gamma_{p+1}(k+1-\bar{c})^p = 0. \quad (13.43)$$

If you restrict either the near or far end of the lag, the number of  $\gamma$  parameters estimated is reduced by one to account for the restriction; if you restrict both the near and far end of the lag, the number of  $\gamma$  parameters is reduced by two.

By default, EViews does not impose constraints.

## How to Estimate Models Containing PDLs

You specify a polynomial distributed lag by the `pdl` term, with the following information in parentheses, each separated by a comma in this order:

- The name of the series
- The lag length (the number of lagged values of the series to be included)
- The degree of the polynomial
- A numerical code to constrain the lag polynomial (optional):

1	constrain the near end of the lag to zero.
2	constrain the far end.
3	constrain both ends.

You may omit the constraint code if you do not want to constrain the lag polynomial. Any number of `pdl` terms may be included in an equation. Each one tells EViews to fit distributed lag coefficients to the series and to constrain the coefficients to lie on a polynomial.

For example,

```
ls sales c pdl(orders,8,3)
```

fits SALES to a constant, and a distributed lag of current and eight lags of ORDERS, where the lag coefficients of ORDERS lie on a 3rd degree polynomial with no endpoint constraints. Similarly,

```
ls div c pdl(rev,12,4,2)
```

fits DIV to a distributed lag of current and 12 lags of REV, where the coefficients of REV lie on a 4th degree polynomial with a constraint at the far end.

The `pdl` specification may also be used in two-stage least squares. If the series in the `pdl` is exogenous, you should include the PDL of the series in the instruments as well. For this purpose,



you may specify `pdl (*)` as an instrument; all `pdl` variables will be used as instruments. For example, if you specify the TSLS equation as

```
sales c inc pdl(orders(-1),12,4)
```

with instruments

```
fed fed(-1) pdl(*)
```

the distributed lag of `ORDERS` will be used as instruments together with `FED` and `FED(-1)`.

Polynomial distributed lags cannot be used in nonlinear specifications.

## Example

The distributed lag model of industrial production (IP) on money (M1) yields the following results:

Dependent Variable: IP				
Method: Least Squares				
Date: 08/15/97 Time: 17:09				
Sample(adjusted): 1960:01 1989:12				
Included observations: 360 after adjusting endpoints				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	40.67568	0.823866	49.37171	0.0000
M1	0.129699	0.214574	0.604449	0.5459
M1(-1)	-0.045962	0.376907	-0.121944	0.9030
M1(-2)	0.033183	0.397099	0.083563	0.9335
M1(-3)	0.010621	0.405861	0.026169	0.9791
M1(-4)	0.031425	0.418805	0.075035	0.9402
M1(-5)	-0.048847	0.431728	-0.113143	0.9100
M1(-6)	0.053880	0.440753	0.122245	0.9028
M1(-7)	-0.015240	0.436123	-0.034944	0.9721
M1(-8)	-0.024902	0.423546	-0.058795	0.9531
M1(-9)	-0.028048	0.413540	-0.067825	0.9460
M1(-10)	0.030806	0.407523	0.075593	0.9398
M1(-11)	0.018509	0.389133	0.047564	0.9621
M1(-12)	-0.057373	0.228826	-0.250728	0.8022
R-squared	0.852398	Mean dependent var	71.72679	
Adjusted R-squared	0.846852	S.D. dependent var	19.53063	
S.E. of regression	7.643137	Akaike info criterion	6.943606	
Sum squared resid	20212.47	Schwarz criterion	7.094732	
Log likelihood	-1235.849	F-statistic	153.7030	
Durbin-Watson stat	0.008255	Prob(F-statistic)	0.000000	

Taken individually, none of the coefficients on lagged M1 is significant. Yet the regression as a whole has a reasonable  $R^2$  with a very significant  $F$ -statistic (though with a very low Durbin-Watson statistic). This is a typical symptom of high collinearity among the regressors and suggests fitting a polynomial distributed lag model.

To estimate a fifth-degree polynomial distributed lag model with no constraints, enter the commands:

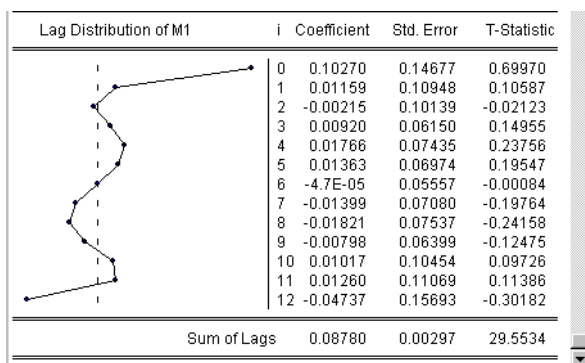
```
smpl 59.1 89.12
ls ip c pdl(m1,12,5)
```

The following result is reported by at the top of the equation window:

Dependent Variable: IP				
Method: Least Squares				
Date: 08/15/97 Time: 17:53				
Sample(adjusted): 1960:01 1989:12				
Included observations: 360 after adjusting endpoints				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	40.67311	0.815195	49.89374	0.0000
PDL01	-4.66E-05	0.055566	-0.000839	0.9993
PDL02	-0.015625	0.062884	-0.248479	0.8039
PDL03	-0.000160	0.013909	-0.011485	0.9908
PDL04	0.001862	0.007700	0.241788	0.8091
PDL05	2.58E-05	0.000408	0.063211	0.9496
PDL06	-4.93E-05	0.000180	-0.273611	0.7845
R-squared	0.852371	Mean dependent var		71.72679
Adjusted R-squared	0.849862	S.D. dependent var		19.53063
S.E. of regression	7.567664	Akaike info criterion		6.904899
Sum squared resid	20216.15	Schwarz criterion		6.980462
Log likelihood	-1235.882	F-statistic		339.6882
Durbin-Watson stat	0.008026	Prob(F-statistic)		0.000000

This portion of the view reports the estimated coefficients  $\gamma$  of the polynomial in Equation (13.38) on page 293. The terms PDL01, PDL02, PDL03, ..., correspond to  $z_1, z_2, \dots$  in Equation (13.40).

The implied coefficients of interest  $\beta_j$  in equation (1) are reported at the bottom of the table, together with a plot of the estimated polynomial:



The Sum of Lags reported at the bottom of the table is the sum of the estimated coefficients on the distributed lag and has the interpretation of the long run effect of M1 on IP, assuming stationarity.

Note that selecting **View/Coefficient Tests** for an equation estimated with PDL terms tests the restrictions on  $\gamma$ , not on  $\beta$ . In this example, the coefficients on the fourth- (PDL05) and fifth-order (PDL06) terms are individually insignificant and very close to zero. To test the joint significance of these two terms, click **View/Coefficient Tests/Wald-Coefficient Restrictions...** and type

$$c(6)=0, \quad c(7)=0$$

in the Wald Test dialog box. (See Chapter 14 for an extensive discussion of Wald tests in EViews.) EViews displays the result of the joint test:

Wald Test:			
Equation: IP_PDL			
Null Hypothesis: C(6)=0			
C(7)=0			
F-statistic	0.039852	Probability	0.960936
Chi-square	0.079704	Probability	0.960932

There is no evidence to reject the null hypothesis, suggesting that you could have fit a lower order polynomial to your lag structure.

## Nonstationary Time Series

The theory behind ARMA estimation is based on stationary time series. A series is said to be (weakly or covariance) *stationary* if the mean and autocovariances of the series do not depend on time. Any series that is not stationary is said to be *nonstationary*.

A common example of a nonstationary series is the *random walk*:

$$y_t = y_{t-1} + \epsilon_t, \quad (13.44)$$

where  $\epsilon$  is a stationary random disturbance term. The series  $y$  has a constant forecast value, conditional on  $t$ , and the variance is increasing over time. The random walk is a difference stationary series since the first difference of  $y$  is stationary:

$$y_t - y_{t-1} = (1 - L)y_t = \epsilon_t. \quad (13.45)$$

A difference stationary series is said to be *integrated* and is denoted as  $I(d)$  where  $d$  is the order of integration. The order of integration is the number of unit roots contained in the series, or the number of differencing operations it takes to make the series stationary. For the random walk above, there is one unit root, so it is an  $I(1)$  series. Similarly, a stationary series is  $I(0)$ .

Standard inference procedures do not apply to regressions which contain an integrated dependent variable or integrated regressors. Therefore, it is important to check whether a series is stationary or not before using it in a regression. The formal method to test the stationarity of a series is the unit root test.

## Unit Root Tests

EViews performs two widely used unit root tests: the Dickey-Fuller (DF) and augmented Dickey-Fuller (ADF) tests, and the Phillips-Perron (PP) test. These tests are available as views of a series. In this section we provide some theoretical background for these two tests; users who only want to know how to implement the test in EViews should skip this section.

### The Augmented Dickey-Fuller (ADF) Test

To illustrate the use of Dickey-Fuller tests, consider first an AR(1) process:

$$y_t = \mu + \rho y_{t-1} + \epsilon_t, \quad (13.46)$$

where  $\mu$  and  $\rho$  are parameters and  $\epsilon_t$  is assumed to be white noise.  $y$  is a stationary series if  $-1 < \rho < 1$ . If  $\rho = 1$ ,  $y$  is a nonstationary series (a random walk with drift); if the process is started at some point, the variance of  $y$  increases steadily with time and goes to infinity. If the absolute value of  $\rho$  is greater than one, the series is explosive. Therefore, the hypothesis of a stationary series can be evaluated by testing whether the absolute value of  $\rho$  is strictly less than one. Both the DF and the PP tests take the unit root as the null hypothesis  $H_0: \rho = 1$ . Since explosive series do not make much economic sense, this null hypothesis is tested against the one-sided alternative  $H_1: \rho < 1$ .

The test is carried out by estimating an equation with  $y_{t-1}$  subtracted from both sides of the equation:

$$\Delta y_t = \mu + \gamma y_{t-1} + \epsilon_t, \quad (13.47)$$

where  $\gamma = \rho - 1$  and the null and alternative hypotheses are

$$\begin{aligned} H_0: \gamma &= 0 \\ H_1: \gamma &< 0 \end{aligned} \quad (13.48)$$

While it may appear that the test can be carried out by performing a  $t$ -test on the estimated  $\gamma$ , the  $t$ -statistic under the null hypothesis of a unit root does not have the conventional  $t$ -distribution. Dickey and Fuller (1979) showed that the distribution under the null hypothesis is non-standard, and simulated the critical values for selected sample sizes. More recently, MacKinnon (1991) has implemented a much larger set of simulations than those tabulated by Dickey and Fuller. In addition, MacKinnon estimates the response surface using the simulation results, permitting the calculation of Dickey-Fuller critical values for any sample size and for any number of right-hand variables. EViews reports these MacKinnon critical values for unit root tests.

The simple unit root test described above is valid only if the series is an AR(1) process. If the series is correlated at higher order lags, the assumption of white noise disturbances is violated. The ADF and PP tests use different methods to control for higher-order serial correlation in the series. The ADF test makes a parametric correction for higher-order correlation by assuming that the  $y$  series follows an AR( $p$ ) process and adjusting the test methodology. The PP approach is described below.

The ADF approach controls for higher-order correlation by adding lagged difference terms of the dependent variable  $y$  to the right-hand side of the regression:

$$\Delta y_t = \mu + \gamma y_{t-1} + \delta_1 \Delta y_{t-1} + \delta_2 \Delta y_{t-2} + \dots + \delta_p \Delta y_{t-p} + \epsilon_t. \quad (13.49)$$

This augmented specification is then used to test:

$$\begin{aligned} H_0: \gamma &= 0 \\ H_1: \gamma &< 0 \end{aligned} \quad (13.50)$$

in this regression. An important result obtained by Fuller is that the asymptotic distribution of the  $t$ -statistic on  $\gamma$  is independent of the number of lagged first differences included in the ADF regression. Moreover, while the parametric assumption that  $y$  follows an autoregressive (AR) process may seem restrictive, Said and Dickey (1984) demonstrate that the ADF test remains valid even when the series has a moving average (MA) component, provided that enough lagged difference terms are augmented to the regression.

EViews refers to both the Dickey-Fuller and the Augmented Dickey-Fuller tests as ADF tests. You will face two practical issues in performing the ADF test. First, you will have to specify the number of lagged first difference terms to add to the test regression (selecting zero yields the DF test; choosing numbers greater than zero generate ADF tests). The usual (though not particularly useful) advice is to include lags sufficient to remove any serial correlation in the residuals.

Second, EViews asks you whether to include other exogenous variables in the test regression. You have the choice of including a constant, a constant and a linear time trend, or neither in the test regression. Your choice here is important since the asymptotic distribution of the  $t$ -statistic under the null hypothesis depends on your assumptions regarding these deterministic terms.

- If you include a constant in the test regression, the  $t$ -statistic has a nonstandard distribution if the underlying process contains a unit root with a zero constant.
- If you include a constant and linear trend in the test regression, the  $t$ -statistic has a nonstandard distribution if the underlying process contains a unit root with a zero linear trend.

The asymptotic distribution changes when these assumptions are not satisfied. For example, if you include a constant in the test regression and if the underlying process contains a unit root with a *nonzero* constant, then the  $t$ -statistic has an asymptotic standard normal distribution under the null hypothesis of a unit root.

Although this result complicates the testing procedure by increasing the number of cases to consider, Hylleberg and Mizon (1989) have shown that the standard normal critical values are likely to lead to too frequent rejection of the null hypothesis even in large samples, unless the constant is very large. They present critical values for selected sample sizes and various positive values of the constant parameter  $\mu$ . Their critical values lie between the Dickey-Fuller values and those from the standard normal distribution; as the value of  $\mu$  becomes smaller, the critical values become closer to the Dickey-Fuller values. Their recommendation is to use the Dickey-Fuller values rather than those from the standard normal distribution for moderate sample sizes.

There still remains the problem of whether to include a constant, a constant and a linear trend, or neither in the test regression. One approach would be to run the test with both a constant and a linear trend since the other two cases are just special cases of this more general specification. However, including irrelevant regressors in the regression reduces the power of the test, possibly concluding that there is a unit root when, in fact, there is none. The general principle is to choose a specification that is a plausible description of the data under both the null and alternative hypotheses (Hamilton 1994, p.501). If the series seems to contain a trend (whether deterministic or stochastic), you should include both a constant and trend in the test regression. If the series does not exhibit any trend and has a nonzero mean, you should only include a constant in the regression, while if the series seems to be fluctuating around a zero mean, you should include neither a constant nor a trend in the test regression.

### The Phillips-Perron (PP) Test

Phillips and Perron (1988) propose a nonparametric method of controlling for higher-order serial correlation in a series. The test regression for the Phillips-Perron (PP) test is the AR(1) process:

$$\Delta y_t = \alpha + \beta y_{t-1} + \epsilon_t \quad (13.51)$$

While the ADF test corrects for higher order serial correlation by adding lagged differenced terms on the right-hand side, the PP test makes a correction to the  $t$ -statistic of the  $\gamma$  coefficient from the AR(1) regression to account for the serial correlation in  $\epsilon$ . The correction is nonparametric since we use an estimate of the spectrum of  $\epsilon$  at frequency zero that is robust to heteroskedasticity and autocorrelation of unknown form. EVIEWS uses the Newey-West heteroskedasticity autocorrelation consistent estimate

$$\hat{\omega}^2 = \gamma_0 + 2 \sum_{v=1}^q \left(1 - \frac{v}{q+1}\right) \gamma_v \quad (13.52)$$

where

$$\gamma_j = \left( \sum_{t=j+1}^T \hat{\epsilon}_t \hat{\epsilon}_{t-j} \right) / T \quad (13.53)$$

and  $q$  is the truncation lag. The PP  $t$ -statistic is computed a

$$t_{pp} = \frac{\gamma_0^{1/2} t_b}{\hat{\omega}} - \frac{(\hat{\omega}^2 - \gamma_0) T s_b}{2 \hat{\omega} s} \quad (13.54)$$

where  $t_b$ ,  $s_b$  are the  $t$ -statistic and standard error of  $\beta$  and  $s$  is the standard error of the test regression.

The asymptotic distribution of the PP  $t$ -statistic is the same as the ADF  $t$ -statistic and EViews again reports MacKinnon critical values. As with the ADF test, you have to specify whether to include a constant, a constant and linear trend, or neither in the test regression. For the PP test, you also have to specify the truncation lag  $q$  for the Newey-West correction, that is, the number of periods of serial correlation to include. The dialog initially contains the Newey-West automatic truncation lag selection (the floor function returns the largest integer not exceeding the argument)

$$q = \text{floor}(4(T/100)^{2/9}) \quad (13.55)$$

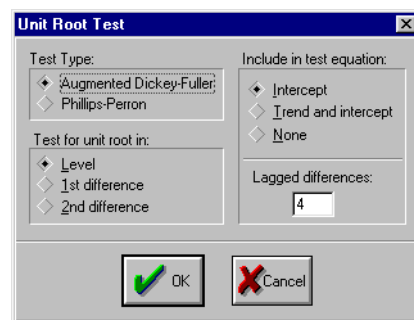
which is based solely on the number of observations used in the test regression. You may, of course, specify any integer.

## Performing Unit Root Tests in EViews

The unit root tests are available as views of a series: double click on the series name to open the series window and choose **View/Unit Root Test...**

You have to specify four things to carry out a unit root test.

1. Choose the type of test, either the Augmented Dickey-Fuller (ADF) test or the Phillips-Perron (PP) test.
2. Specify whether to test for a unit root in the level, first difference, or second difference of the series. You can use this option to determine the number of unit roots in the series. If the test fails to reject the test in levels but rejects the test in first differences, then the series contains one unit root and is of integrated order one  $I(1)$ . If the test fails to reject the test in levels and first differences but rejects the test in second differences, then the series contains two unit roots and is of integrated order two  $I(2)$ .
3. Specify whether to include a constant, a constant and linear trend, or neither in the test regression. The choice is important since the distribution of the test statistic under the null hypothesis differs among these three cases.
4. Select the order of serial correlation to account for in the series. For the ADF test, you



specify the number of lagged first difference terms to add in the test regression. For the PP test, you specify the lag truncation to compute the Newey-West heteroskedasticity and autocorrelation (HAC) consistent estimate of the spectrum at zero frequency.

Having specified these options, click **OK** to carry out the test. EViews reports the test statistic together with the estimated test regression:

ADF Test Statistic	-0.223277	1% Critical Value*	-3.4502
		5% Critical Value	-2.8696
		10% Critical Value	-2.5711

\*MacKinnon critical values for rejection of hypothesis of a unit root.

Augmented Dickey-Fuller Test Equation

Dependent Variable: D(IP)

Method: Least Squares

Date: 10/17/97 Time: 10:44

Sample(adjusted): 1959:06 1989:12

Included observations: 367 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
IP(-1)	-0.000343	0.001536	-0.223277	0.8234
D(IP(-1))	0.319733	0.052689	6.068348	0.0000
D(IP(-2))	0.091687	0.055219	1.660416	0.0977
D(IP(-3))	0.050756	0.055258	0.918521	0.3590
D(IP(-4))	0.009248	0.052766	0.175264	0.8610
C	0.121243	0.113067	1.072306	0.2843
R-squared	0.147293	Mean dependent var	0.183815	
Adjusted R-squared	0.135483	S.D. dependent var	0.627660	
S.E. of regression	0.583595	Akaike info criterion	1.776993	
Sum squared resid	122.9503	Schwarz criterion	1.840841	
Log likelihood	-320.0783	F-statistic	12.47157	
Durbin-Watson stat	1.993587	Prob(F-statistic)	0.000000	

For the ADF test, the test statistic is the  $t$ -statistic for the lagged dependent variable in the test regression reported at the bottom part of the table (not shown here). For the PP test, the test statistic is a modified  $t$ -statistic as described above.

The null hypothesis of a unit root is rejected against the one-sided alternative if the  $t$ -statistic is less than (lies to the left of) the critical value. In this example, the test fails to reject the null hypothesis of a unit root in the IP series at any of the reported significance levels.

You should note that the critical values reported by EViews are valid only for unit root tests of a data series, and will be invalid if the series is based on estimated values. For example, Engle and Granger (1987) proposed a two-step method to test for cointegration. The test amounts to testing for a unit root in the residuals of a first stage regression. Since these residuals are estimates of the disturbance term, the asymptotic distribution of the test statistic differs from the one for ordinary series. The critical values reported by EViews are not appropriate for unit root tests in the



residuals of a regression; the correct critical values can be found in Davidson and MacKinnon (1993, Table 20.2).

After running a unit root test, you should examine the estimated test regression reported by EViews, especially if you are not sure about the lag structure or deterministic trend in the series. You may want to re-run the test equation with a different selection of right-hand variables (add or delete the constant, trend, or lagged differences) or lag order.

## Commands

The command

```
equation eq_gdp.ls gdp c ar(1) ar(2) ma(1) ma(2)
```

fits an ARMA(2,2) model to the GDP series and stores the results in the equation object named EQ\_GDP.

```
eq1.auto(4)
```

tests for serial correlation of up to order four in the residuals from equation EQ1.

```
eq1.correlogram(12)
```

displays the correlogram for the residuals in EQ1 up to lag 12.

```
equation eq2.ls gdp c pdl(m1,12,3)
```

fits a third degree polynomial to the coefficients of M1 up to twelve lags.

```
gdp.uroot(4,c)
```

runs the ADF unit root test including a constant and four lags of first differences.

```
gdp.uroot(4,t,h)
```

runs the Phillips-Perron unit root test including a constant and linear trend truncating the Bartlett kernel at lag four.



## Chapter 14. Forecasting from an Equation

---

This chapter describes procedures for forecasting and computing fitted values from a single equation. The techniques described here are for forecasting with equation objects estimated using regression methods. Forecasts from equations estimated by specialized techniques, such as ARCH, binary, ordered, tobit, and count methods, are discussed in the corresponding chapters. Forecasting from a series using exponential smoothing methods is explained in Chapter 7, and forecasting using multiple equations and models is described in Chapter 23.

### Forecasting from Equations in EViews

To illustrate the process of forecasting from an estimated equation, we begin with a simple example. Suppose we have data on the logarithm of monthly housing starts (HS) and the logarithm of the S&P index (SP) over the period 1959:01–1996:01. The data are contained in a workfile with range 1959:01–1998:12.

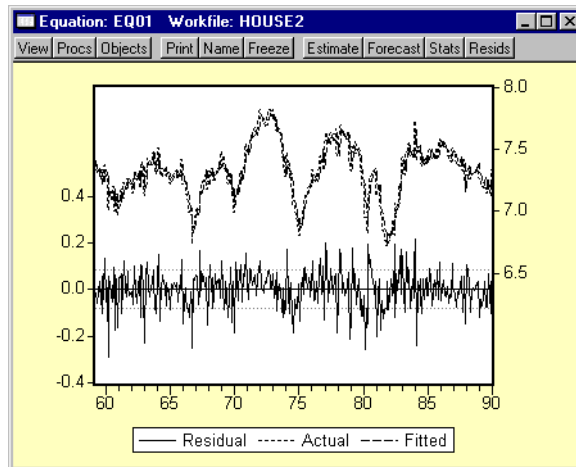
We estimate a regression of HS on a constant, SP, and the lag of HS, with an AR(1) to correct for residual serial correlation, using data for the period 1959:01–1990:12, and then use the model to forecast housing starts under a variety of settings. Following estimation, the equation results are held in the equation object EQ01:

Dependent Variable: HS  
Method: Least Squares  
Date: 10/18/97 Time: 12:44  
Sample(adjusted): 1959:03 1990:01  
Included observations: 371 after adjusting endpoints  
Convergence achieved after 4 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.321924	0.117278	2.744975	0.0063
HS(-1)	0.952653	0.016218	58.74157	0.0000
SP	0.005222	0.007588	0.688249	0.4917
AR(1)	-0.271254	0.052114	-5.205027	0.0000
R-squared	0.861373	Mean dependent var	7.324051	
Adjusted R-squared	0.860240	S.D. dependent var	0.220996	
S.E. of regression	0.082618	Akaike info criterion	-2.138453	
Sum squared resid	2.505050	Schwarz criterion	-2.096230	
Log likelihood	400.6830	F-statistic	760.1338	
Durbin-Watson stat	2.013460	Prob(F-statistic)	0.000000	
Inverted AR Roots	-.27			

Note that the estimation sample is adjusted by two observations to account for the first difference of the lagged endogenous variable used in deriving AR(1) estimates for this model.

To get a feel for the fit of the model, select **View/Actual, Fitted, Residual...**, then choose **Actual, Fitted, Residual Graph**:



The actual and fitted values depicted on the upper portion of the graph are virtually indistinguishable. This view provides little control over the process of producing fitted values, and does not allow you to save your fitted values. These limitations are overcome by using EViews built-in forecasting procedures to compute fitted values for the dependent variable.

### How to Perform a Forecast

To forecast HS from this equation, push the **Forecast** button on the equation toolbar, or select **Procs/Forecast...**

You should provide the following information.

- **Series names**
- **Forecasted series.** Fill in the edit box with the name to be given to the forecasted dependent variable. EViews suggests a name, but you can change it to any valid series name. The name should be different from the name of the dependent variable, since the forecast procedure will overwrite the data in the specified series.
- **S.E. (optional).** If desired, you may provide a name for the series to be filled with the forecast standard errors. If you do not provide a name, no forecast errors will be saved.
- **GARCH (optional).** For models estimated by ARCH, you will be given a further option of saving forecasts of the conditional variances (GARCH terms). See Chapter 16 for a discussion of GARCH estimation.
- **Forecasting method.** You have a choice between the following methods:

**Dynamic**—calculates *multi-step forecasts* starting from the first period in the forecast sample.

**Static**—calculates a sequence of *one-step-ahead forecasts*, using actual, rather than forecasted values for lagged dependent variables.

and you can set the following option:

**Structural**—instructs EViews to ignore any ARMA terms in the equation when forecasting. By default, when your equation has ARMA terms, both dynamic and static solution methods form forecasts of the residuals. If you select **Structural**, all forecasts will ignore the residuals and will form predictions using only the structural part of the model.

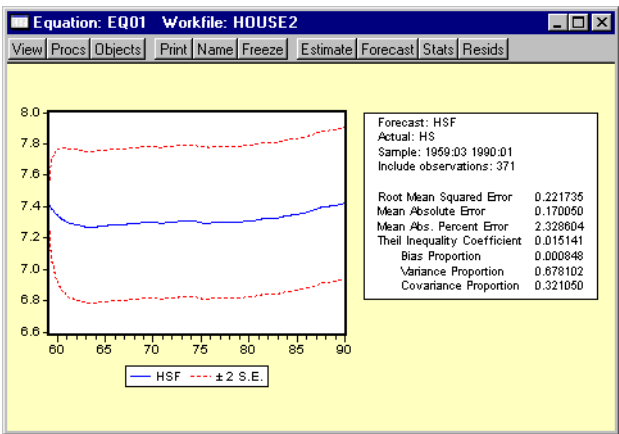
**Sample range.** You must specify the sample to be used for the forecast. By default, EViews sets this sample to be the workfile sample. By specifying a sample outside the sample used in estimating your equation (the *estimation sample*), you can instruct EViews to produce out-of-sample forecasts.

Note that you are responsible for supplying the values for the independent variables in the out-of-sample forecasting period. For static forecasts, you must also supply the values for any lagged dependent variables.

- **Output.** You can choose to see the forecast output as a graph or a numerical forecast evaluation, or both. Forecast evaluation is only available if the forecast sample includes observations for which the dependent variable is observed.

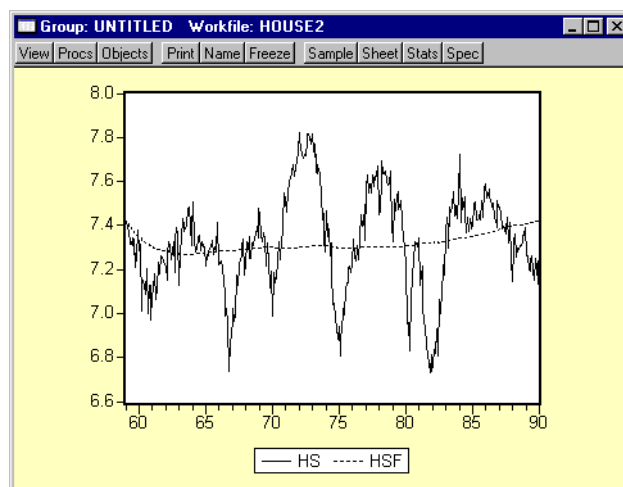
Illustration

Suppose we produce a dynamic forecast using EQ01 over the sample 1959:01 to 1996:01. The forecast values will be placed in the series HSF, and EViews will display both a graph of the forecasts and the plus and minus two standard error bands, as well as a forecast evaluation:



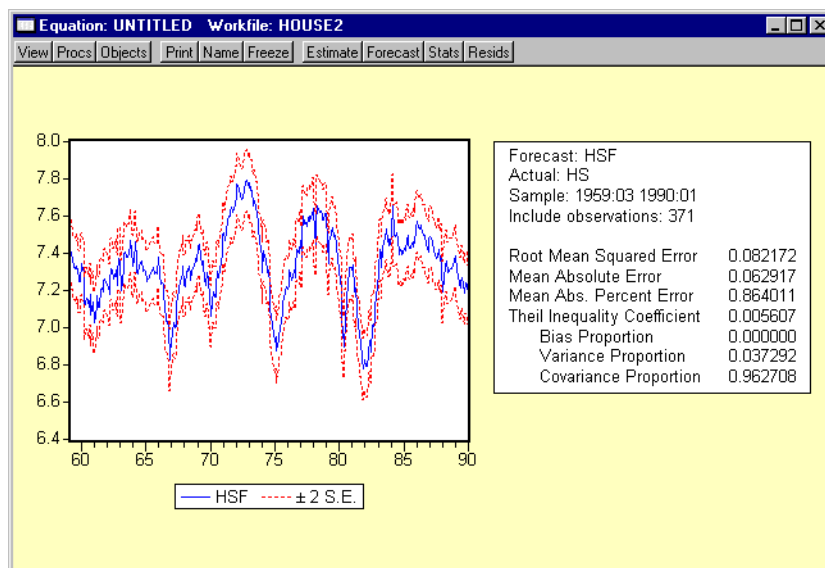
As noted in the output, the forecast values are saved in the series HSF. Since HSF is a standard EViews series, you may examine your forecasts using all of the standard tools for working with series objects.

We can examine the actual versus fitted values by creating a group containing HS and HSF, and plotting the two series. Click on **Quick/Show...** and enter HS and HSF. Then select **View/Graph/Line** to display the two series:

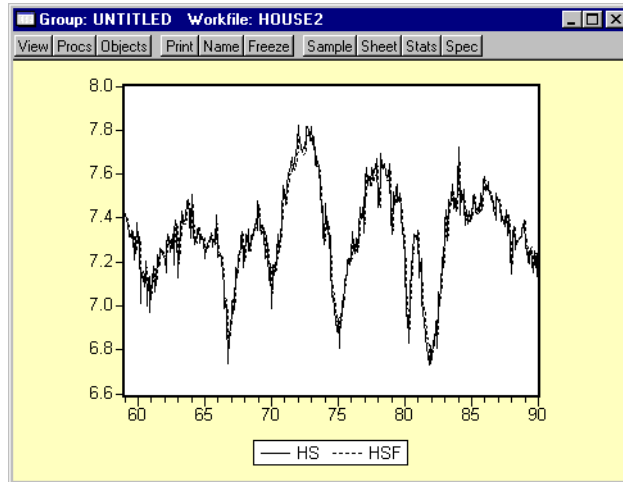


This is a dynamic forecast over the entire period from 1959:01 through 1996:01. For every period, the previously forecasted values for HS(-1) are used in forming a forecast of the subsequent value of HS. Note the considerable difference between this actual and fitted graph and the **Actual, Fitted, Residual Graph** depicted above.

To perform a series of one-step ahead forecasts, click on **Forecast** on the equation toolbar, and select **Static** forecasts. EViews will display the forecast results:



We can also compare the actual and fitted values from the static forecast by examining a line graph of a group containing HS and HSF:



The one-step-ahead static forecasts are more accurate than the dynamic forecasts since, for each period, the actual value of  $HS(-1)$  is used in forming the forecast of HS. These one-step-ahead static forecasts are the same forecasts used in the **Actual, Fitted, Residual Graph** displayed above.

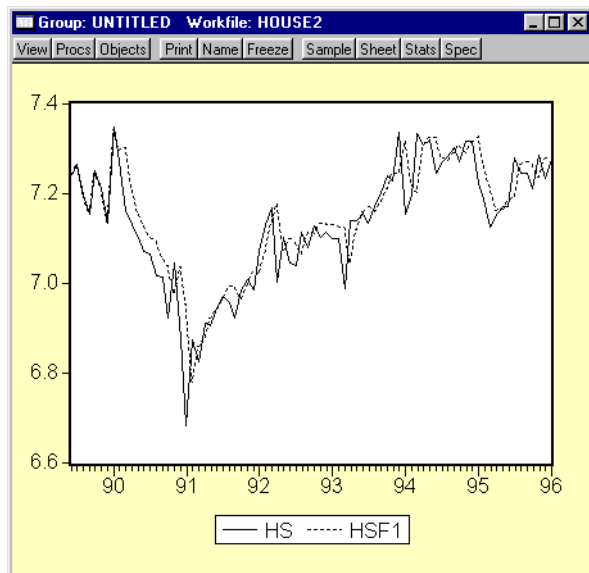
Lastly, we construct a dynamic forecast beginning in 1990:02 (the first period following the estimation sample) and ending in 1996:01. Keep in mind that data are available for SP for this entire period. The plot of the actual and the forecast values for 1989:01 to 1996:01 is given by:





EViews backfills the forecast series prior to the forecast sample (up through 1990:01) and then dynamically forecasts HS for each subsequent period through 1996:01. This is the forecast that you would have constructed if, in 1990:01, you predicted values of HS from 1990:02 through 1996:01, given knowledge about the entire path of SP over that period.

The corresponding static forecast is



Again, EViews backfills the values of the forecast series, HSF1, through 1990:01. This forecast is the one you would have constructed if, in 1990:01, you used all available data to estimate a model, and then used this estimated model to perform one-step ahead forecasts every month for the next six years.

The remainder of this chapter focuses on the details associated with the construction of these forecasts, and the corresponding forecast evaluations.

## Forecast Basics

EViews stores the forecast results in the series specified in the **Forecast name** field. We will refer to this series as the *forecast series*.

The *forecast sample* specifies the observations for which EViews will try to compute fitted or forecasted values. If the forecast is not computable, a missing value will be returned. In some cases, EViews will carry out automatic adjustment of the sample to prevent a forecast consisting entirely of missing values (see “Adjustment for missing values” on page 313, below). Note that the forecast sample may or may not overlap with the sample of observations used to estimate the equation.

For values not included in the forecast sample, there are two options. By default, EViews fills in the actual values of the dependent variable. If you turn off the **Insert actuals for out-of-sample** option, out-of-forecast-sample values will be filled with NAs.

As a consequence of these rules, *all data in the forecast series will be overwritten during the forecast procedure*. Existing values in the forecast series will be lost.

## Computing forecasts

For each observation in the forecast sample, EViews computes the fitted value of the dependent variable using the estimated parameters, the right-hand side exogenous variables, and, either the actual or estimated values for lagged endogenous variables and residuals. The method of constructing these forecasted values depends upon the estimated model and user specified settings.

To illustrate the forecasting procedure, we begin with a simple linear regression model with no lagged endogenous right-hand side variables, and no ARMA terms. Suppose that you have estimated the following equation specification:

y c x z

Now click on **Forecast**, specify a forecast period, and click **OK**.

For every observation in the forecast period, EViews will compute the fitted value of Y using the estimated parameters and the corresponding values of the regressors, X and Z:

$$\hat{y}_t = \hat{c}(1) + \hat{c}(2)x_t + \hat{c}(3)z_t. \quad (14.1)$$

You should make certain that you have valid values for the exogenous right-hand side variables for all observations in the forecast period. If any data are missing in the forecast sample, the corresponding forecast observation will be an NA.

### Adjustment for missing values

There are two cases when a missing value will be returned for the forecast value. First, if any of the regressors have a missing value, and second, if any of the regressors are out of the range of the workfile. This includes the implicit error terms in AR models.

In the case of forecasts with no dynamic components in the specification (*i.e.* with no lagged endogenous or ARMA error terms), a missing value in the forecast series will not affect subsequent forecasted values. In the case where there are dynamic components, however, a single missing value in the forecasted series will propagate throughout all future values of the series.

As a convenience feature, EViews will move the starting point of the sample forward where necessary, until a valid forecast value is obtained. Without these adjustments, the user would have to figure out the appropriate number of presample values to skip, otherwise the forecast would consist entirely of missing values. For example, suppose you wanted to forecast dynamically from the following equation specification:

y c y(-1) ar(1)

If you specified the beginning of the forecast sample to the beginning of the workfile range, EViews will adjust forward the forecast sample by 2 observations, and will use the pre-forecast-sample values of the lagged variables (the loss of 2 observations occurs because the residual loses one observation due to the lagged endogenous variable so that the forecast for the error term can begin only from the third observation.)

### Forecast Errors and Variances

Suppose the “true” model is given by:

$$y_t = x_t' \beta + \epsilon_t, \quad (14.2)$$

where  $\epsilon_t$  is an independent, and identically distributed, mean zero random disturbance, and  $\beta$  is a vector of unknown parameters. Below, we relax the restriction that the  $\epsilon$ ’s be independent.

The true model generating  $y$  is not known, but we obtain estimates  $b$  of the unknown parameters  $\beta$ . Then, setting the error term equal to its mean value of zero, the (point) forecasts of  $y$  are obtained as

$$\hat{y}_t = x_t' b. \quad (14.3)$$

Forecasts are made with error, where the error is simply the difference between the actual and forecasted value  $e_t = y_t - x_t' b$ . Assuming that the model is correctly specified, there are two sources of forecast error: residual uncertainty and coefficient uncertainty.

### Residual Uncertainty

The first source of error, termed *residual* or *innovation uncertainty*, arises because the innovations  $\epsilon$  in the equation are unknown for the forecast period, and are replaced with their expectations. While the residuals are zero in expected value, the individual values are non-zero; the larger the variation in the individual errors, the greater the overall error in the forecasts.

The standard measure of this variation is the standard error of the regression (labeled “S.E. of regression” in the equation output). Residual uncertainty is usually the largest source of forecast error.

In dynamic forecasts, innovation uncertainty is compounded by the fact that lagged dependent variables and ARMA terms depend on lagged innovations. EViews also sets these equal to their expected values, which differ randomly from realized values. This additional source of forecast uncertainty tends to rise over the forecast horizon, leading to a pattern of increasing forecast errors. Forecasting with lagged dependent variables and ARMA terms is discussed in more detail below.

### Coefficient Uncertainty

The second source of forecast error is *coefficient uncertainty*. The estimated coefficients  $b$  of the equation deviate from the true coefficients  $\beta$  in a random fashion. The standard error of the estimated coefficient, given in the regression output, is a measure of the precision with which the estimated coefficients measure the true coefficients.

The effect of coefficient uncertainty depends upon the exogenous variables. Since the estimated coefficients are multiplied by the exogenous variables  $x$  in the computation of forecasts, the more the exogenous variables deviate from their mean values, the greater is the forecast uncertainty.

### Forecast Variability

The variability of forecasts is measured by the forecast standard errors. For a single equation without lagged dependent variables or ARMA terms, the forecast standard errors are computed as

$$\text{forecast se} = s \sqrt{1 + x_t'(X'X)x_t}. \quad (14.4)$$

where  $s$  is the standard error of regression. These standard errors account for both innovation (the first term) and coefficient uncertainty (the second term). Point forecasts made from linear regression models estimated by least squares are optimal in the sense that they have the smallest forecast variance among forecasts made by linear unbiased estimators. Moreover, if the innovations are normally distributed, the forecast errors have a  $t$ -distribution and forecast intervals can be readily formed.

If you supply a name for the forecast standard errors, EViews computes and saves a series of forecast standard errors in your workfile. You can use these standard errors to form forecast intervals. If you choose the **Do graph** option for output, EViews will plot the forecasts with plus and minus two standard error bands. These two standard error bands provide an approximate 95% forecast interval; if you (hypothetically) make many forecasts, the actual value of the dependent variable will fall inside these bounds 95 percent of the time.

### Additional Details

EViews accounts for the additional forecast uncertainty generated when lagged dependent variables are used as explanatory variables (see “Forecasts with Lagged Dependent Variables” on page 318).

There are several other special cases, involving dependent variables that are defined by expression, where coefficient uncertainty is ignored. These cases are described later in this chapter (“Forecasting Equations with Formulas” on page 323).

Forecast standard errors derived from equations estimated by nonlinear least squares and equations that include PDL (polynomial distributed lag) terms only account for the residual uncertainty (“Forecasting with Nonlinear and PDL Specifications” on page 327).

### Forecast Evaluation

Suppose we construct a dynamic forecast for HS over the period 1990:02 to 1996:01 using our estimated housing equation. If the **Forecast evaluation** option is checked, and there are actual data for the forecasted variable for the forecast sample, EViews reports a table of statistical results evaluating the forecast:

Forecast: HSF	
Actual: HS	
Sample: 1990:02 1996:01	
<u>Include observations: 72</u>	
Root Mean Squared Error	0.318700
Mean Absolute Error	0.297261
Mean Absolute Percentage Error	4.205889
Theil Inequality Coefficient	0.021917
Bias Proportion	0.869982
Variance Proportion	0.082804
Covariance Proportion	0.047214

Note that EViews cannot compute a forecast evaluation if there are no data for the dependent variable for the forecast sample.

The forecast evaluation is saved in one of two formats. If you turn on the **Do graph** option, the forecasts are included along with a graph of the forecasts. If you wish to display the evaluations in their own table, you should turn off the **Do graph** option in the Forecast dialog box.

Suppose the forecast sample is  $t = T + 1, \dots, T + h$  and denote the actual and forecasted value in period  $t$  as  $y_t$  and  $\hat{y}_t$ , respectively. The reported forecast error statistics are computed as follows:

Root Mean Squared Error	$\sqrt{\sum_{t=T+1}^{T+h} (\hat{y}_t - y_t)^2 / (h + 1)}$
Mean Absolute Error	$\sum_{t=T+1}^{T+h}  \hat{y}_t - y_t  / (h + 1)$
Mean Absolute Percentage Error	$\sum_{t=T+1}^{T+h} \left  \frac{\hat{y}_t - y_t}{y_t} \right  / (h + 1)$
Theil Inequality Coefficient	$\frac{\sqrt{\sum_{t=T+1}^{T+h} (\hat{y}_t - y_t)^2 / (h + 1)}}{\sqrt{\sum_{t=T+1}^{T+h} \hat{y}_t^2 / (h + 1)} + \sqrt{\sum_{t=T+1}^{T+h} y_t^2 / (h + 1)}}$

The first two forecast error statistics depend on the scale of the dependent variable. These should be used as relative measures to compare forecasts for the same series across different models; the smaller the error, the better the forecasting ability of that model according to that

criterion. The remaining two statistics are scale invariant. The Theil inequality coefficient always lies between zero and one, where zero indicates a perfect fit.

The mean squared forecast error can be decomposed as

$$\sum (\hat{y}_t - y_t)^2 / h = ((\sum \hat{y}_t / h) - \bar{y})^2 + (s_{\hat{y}} - s_y)^2 + 2(1 - r)s_{\hat{y}}s_y \quad (14.5)$$

where  $\sum \hat{y}_t / h$ ,  $\bar{y}$ ,  $s_{\hat{y}}$ ,  $s_y$  are the means and (biased) standard deviations of  $\hat{y}_t$  and  $y_t$ , and  $r$  is the correlation between  $\hat{y}_t$  and  $y_t$ . The proportions are defined as:

Bias Proportion	$\frac{((\sum \hat{y}_t / h) - \bar{y})^2}{\sum (\hat{y}_t - y_t)^2 / h}$
Variance Proportion	$\frac{(s_{\hat{y}} - s_y)^2}{\sum (\hat{y}_t - y_t)^2 / h}$
Covariance Proportion	$\frac{2(1 - r)s_{\hat{y}}s_y}{\sum (\hat{y}_t - y_t)^2 / h}$

- The bias proportion tells us how far the mean of the forecast is from the mean of the actual series.
- The variance proportion tells us how far the variation of the forecast is from the variation of the actual series.
- The covariance proportion measures the remaining unsystematic forecasting errors.

Note that the bias, variance, and covariance proportions add up to one.

If your forecast is “good”, the bias and variance proportions should be small so that most of the bias should be concentrated on the covariance proportions. For additional discussion of forecast evaluation, see Pindyck and Rubinfeld (1991, Chapter 12).

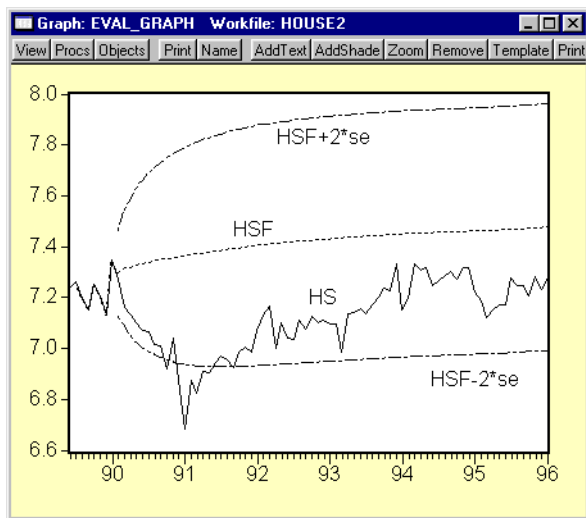
For the example output, the bias proportion is large, indicating that the mean of the forecasts does a poor job of tracking the mean of the dependent variable. To check this, we will plot the forecasted series together with the actual series in the forecast sample with the two standard error bounds. Suppose we saved the forecasts and their standard errors as HSF and HSFSE, respectively. Then the plus and minus two standard error series can be generated by the commands

```

smpl 1990:02 1996:01
series hsf_high = hsf + 2*hsfse
series hsf_low = hsf - 2*hsfse

```

Create a group containing the four series. You can highlight the four series HS, HSF, HSF\_HIGH, and HSF\_LOW, double click on the selected area, and select **Open Group**, or you can select **Quick/Show...** and enter the four series names. Once you have the group open, select **View/Graph/Line**.



The forecasts completely miss the downturn at the start of the 1990's, but, subsequent to the recovery, track the trend reasonably well from 1992 to 1996.

## Forecasts with Lagged Dependent Variables

Forecasting is complicated by the presence of lagged dependent variables on the right-hand side of the equation. For example, we can augment the earlier specification to include the first lag of Y:

$$y = c + xz + y(-1)$$

and click on the **Forecast** button and fill out the series names in the dialog as above.

There is some question, however, as to how we should evaluate the lagged value of Y that appears on the right-hand side of the equation. There are two possibilities: dynamic forecasting and static forecasting.

## Dynamic Forecasting

If you select dynamic forecasting, EViews will perform a multi-step forecast of Y, beginning at the start of the forecast sample. For our single lag specification above:

- The initial observation in the forecast sample will use the actual value of lagged Y. Thus, if S is the first observation in the forecast sample, EViews will compute

$$\hat{y}_S = \hat{c}(1) + \hat{c}(2)x_S + \hat{c}(3)z_S + \hat{c}(4)y_{S-1}, \quad (14.6)$$

- where  $y_{S-1}$  is the value of the lagged endogenous variable in the period prior to the start of the forecast sample. This is the one-step ahead forecast.
- Forecasts for subsequent observations will use the previously *forecasted* values of Y:



$$\hat{y}_{S+k} = \hat{c}(1) + \hat{c}(2)x_{S+k} + \hat{c}(3)z_{S+k} + \hat{c}(4)\hat{y}_{S+k-1}. \quad (14.7)$$

- These forecasts may differ significantly from the one-step ahead forecasts.

If there are additional lags of  $Y$  in the estimating equation, the above algorithm is modified to account for the non-availability of lagged forecasted values in the additional period. For example, if there are three lags of  $Y$  in the equation:

- The first observation (  $S$  ) uses the actual values for all three lags,  $y_{S-3}$ ,  $y_{S-2}$ , and  $y_{S-1}$ .
- The second observation (  $S+1$  ) uses actual values for  $y_{S-2}$  and  $y_{S-1}$  and the forecasted value  $\hat{y}_S$  of the first lag of  $y_{S+1}$ .
- The third observation (  $S+2$  ) will use the actual values for  $y_{S-1}$ , and forecasted values  $\hat{y}_{S+1}$  and  $\hat{y}_S$  for the first and second lags of  $y_{S+2}$ .
- All subsequent observations will use the forecasted values for all three lags.

The selection of the start of the forecast sample is very important for dynamic forecasting. The dynamic forecasts are true multi-step forecasts (from the start of the forecast sample), since they use the recursively computed forecast of the lagged value of the dependent variable. These forecasts may be interpreted as the forecasts for subsequent periods that would be computed using information available at the start of the forecast sample.

Dynamic forecasting requires that data for the exogenous variables be available for every observation in the forecast sample, and that values for any lagged dependent variables be observed at the start of the forecast sample (in our example,  $y_{S-1}$ , but more generally, any lags of  $y$ ). If necessary, the forecast sample will be adjusted.

Any missing values for the explanatory variables will generate an NA for that observation and in all subsequent observations, via the dynamic forecasts of the lagged dependent variable.

## Static Forecasting

Static forecasting performs a series of one-step ahead forecasts of the dependent variable:

- For each observation in the forecast sample, EViews computes

$$\hat{y}_{S+k} = \hat{c}(1) + \hat{c}(2)x_{S+k} + \hat{c}(3)z_{S+k} + \hat{c}(4)y_{S+k-1} \quad (14.8)$$

always using the actual value of the lagged endogenous variable. (14.9)

Static forecasting requires that data for both the exogenous and any lagged endogenous variables be observed for every observation in the forecast sample. As above, EViews will, if necessary, adjust the forecast sample to account for pre-sample lagged variables. If the data are not available for any period, the forecasted value for that observation will be an NA. The presence of a forecasted value of NA does not have any impact on forecasts for subsequent observations.

## A Comparison of Dynamic and Static Forecasting

Both methods will always yield identical results in the first period of a multi-period forecast. Thus, two forecast series, one dynamic and the other static, should be identical for the first observation in the forecast sample.

The two methods will differ for subsequent periods only if there are lagged dependent variables or ARMA terms.

## Forecasting with ARMA errors

Forecasting from equations with ARMA components involves some additional complexities. When you use the AR or MA specifications, you will need to be aware of how EViews handles the forecasts of the lagged residuals which are used in forecasting.

### Structural Forecasts

By default, EViews will forecast values for the residuals using the estimated ARMA structure, as described below.

For some types of work, you may wish to assume that the ARMA errors are always zero. If you select the structural forecast option by checking **Structural (ignore ARMA)**, EViews computes the forecasts assuming that the errors are always zero. If the equation is estimated without ARMA terms, this option has no effect on the forecasts.

### Forecasting with AR errors

For equations with AR errors, EViews adds forecasts of the residuals from the equation to the forecast of the structural model that is based on the right-hand side variables.

In order to compute an estimate of the residual, EViews requires estimates or actual values of the lagged residuals. For the first observation in the forecast sample, EViews will use pre-sample data to compute the lagged residuals. If the pre-sample data needed to compute the lagged residuals are not available, EViews will adjust the forecast sample, and backfill the forecast series with actual values (see the discussion of “Adjustment for missing values” on page 313, above).

If you choose the **Dynamic** option, both the lagged dependent variable and the lagged residuals will be forecasted dynamically. If you select **Static**, both will be set to the actual lagged values. For example, consider the following AR(2) model:

$$\begin{aligned}y_t &= x_t' \beta + u_t \\u_t &= \rho_1 u_{t-1} + \rho_2 u_{t-2} + \epsilon_t\end{aligned}\tag{14.10}$$

Denote the fitted residuals as  $e_t = y_t - x_t'b$ , and suppose the model was estimated using data up to  $t = S - 1$ . Then, provided that the  $x_t$  values are available, the static and dynamic forecasts for  $t = S, S + 1, \dots$ , are given by:

	static	dynamic
$\hat{y}_S$	$x_S'b + \hat{\rho}_1 e_{S-1} + \hat{\rho}_2 e_{S-2}$	$x_S'b + \hat{\rho}_1 e_{S-1} + \hat{\rho}_2 e_{S-2}$
$\hat{y}_{S+1}$	$x_{S+1}'b + \hat{\rho}_1 e_S + \hat{\rho}_2 e_{S-1}$	$x_{S+1}'b + \hat{\rho}_1 \hat{u}_S + \hat{\rho}_2 e_{S-1}$
$\hat{y}_{S+2}$	$x_{S+2}'b + \hat{\rho}_1 e_{S+1} + \hat{\rho}_2 e_S$	$x_{S+2}'b + \hat{\rho}_1 \hat{u}_{S+1} + \hat{\rho}_2 \hat{u}_S$

where the residuals  $\hat{u}_t = \hat{y}_t - x_t'b$  are formed using the forecasted values of  $y_t$ . For subsequent observations, the dynamic forecast will always use the residuals based upon the multi-step forecasts, while the static forecast will use the one-step ahead forecast residuals.

## Forecasting with MA errors

In general, you need not concern yourselves with the details of MA forecasting, since EViews will do all of the work for you. For those of you who are interested in the details of dynamic forecasting, however, the following discussion should aid you in relating EViews results with those obtained from other sources.

The first step in computing forecasts using MA terms is to obtain fitted values for the innovations in the pre-forecast sample period. For example, if you are forecasting the values of  $y$ , beginning in period  $S$ , with a simple MA( $q$ ),

$$\hat{y}_S = \epsilon_S + \hat{\phi}_1 \epsilon_{S-1} + \dots + \hat{\phi}_q \epsilon_{S-q}, \quad (14.11)$$

you will need values for the lagged innovations,  $\epsilon_{S-1}, \epsilon_{S-2}, \dots, \epsilon_{S-q}$ .

To compute these pre-forecast innovations, EViews will first assign values for the  $q$  innovations prior to the start of the *estimation* sample,  $\epsilon_0, \epsilon_{-1}, \epsilon_{-2}, \dots, \epsilon_{-q}$ . If your equation is estimated with backcasting turned on, EViews will perform backcasting to obtain these values. If your equation is estimated with backcasting turned off, or if the forecast sample precedes the estimation sample, the initial values will be set to zero.

Given the initial values, EViews will fit the values of subsequent innovations,  $\epsilon_1, \epsilon_2, \dots, \epsilon_q, \dots, \epsilon_{S-1}$ , using forward recursion. The backcasting and recursion procedures are described in detail in the discussion of backcasting in ARMA models in Chapter 12, page 290.

Note the difference between this procedure and the approach for AR errors outlined above, in which the forecast sample is adjusted forward and the pre-forecast values are set to actual values.

The choice between dynamic and static forecasting has two primary implications:

- Once the  $q$  pre-sample values for the innovations are computed, dynamic forecasting sets subsequent innovations to zero. Static forecasting extends the forward recursion through the end of the estimation sample, allowing for a series of one-step ahead forecasts of both the structural model and the innovations.
- When computing static forecasts, EViews uses the entire estimation sample to backcast the innovations. For dynamic MA forecasting, the backcasting procedure uses observations from the beginning of the estimation sample to either the beginning of the forecast period, or the end of the estimation sample, whichever comes first.

### Example

As an example of forecasting from ARMA models, consider forecasting the monthly new housing starts (HS) series. The estimation period is 1959:01–1984:12 and we forecast for the period 1985:01–1991:12. We estimated the following simple multiplicative seasonal autoregressive model:

```
hs c ar(1) sar(12)
```

To forecast from this estimated model, click **Forecast** on the equation toolbar. The forecast evaluation statistics for the model are shown below:

Forecast Evaluation	
Forecast: HS_F	
Actual: HS	
Sample: 1985:01 1991:12	
Include observations: 84	
Root Mean Squared Error	29.13365
Mean Absolute Error	23.35692
Mean Absolute Percentage Error	24.22213
Theil Inequality Coefficient	0.113026
Bias Proportion	0.106193
Variance Proportion	0.609473
Covariance Proportion	0.284335

The large variance proportion indicates that the forecasts are not tracking the variation in the actual HS series. To plot the actual and forecasted series together with the two standard error bands, you can type

```
smpl 1985:01 1991:12
plot hs hs_f hs_f+2*hs_se hs_f-2*hs_se
```

where HS\_F and HS\_SE are the forecasts and standard errors of HS.

As indicated by the large variance proportion, the forecasts track the seasonal movements in HS only at the beginning of the forecast sample and quickly flattens out to the mean forecast value.

## Forecasting Equations with Formulas

EViews allows estimation and forecasting with equations where the left-hand variable is a transformation specified by a formula. When forecasting from equations with formulas on the left-hand side, three things determine the forecasting procedures and options that are available:

- whether the formula is linear or nonlinear
- whether the formula includes lagged variables
- whether the formula includes estimated coefficients

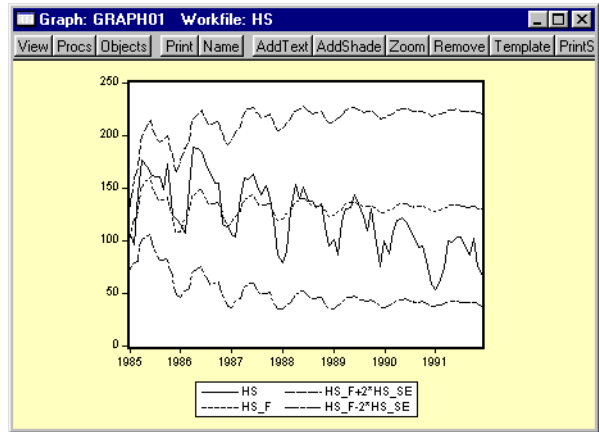
### Point Forecasts

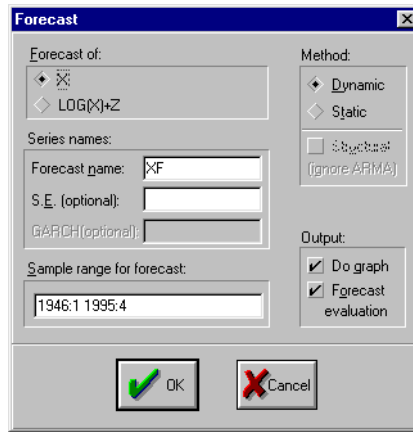
EViews always provides you with the option to forecast the transformed dependent variable. If the transformation can be normalized and solved for the first series in the formula, then EViews also provides you with the option to forecast the normalized series.

For example, suppose you estimated an equation with the specification

$$(\log(x) + z) \subset y$$

If you press the **Forecast** button, the Forecast dialog looks like this:



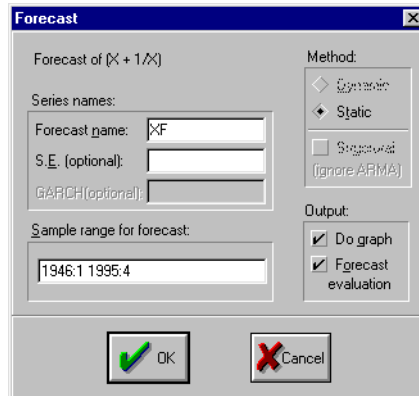


Notice that the dialog provides you with two choices for the series to forecast: the normalized series,  $X$ , and the dependent variable,  $\text{LOG}(X) + Z$ .  $X$  is the normalized series since it is the first series that appears on the left-hand side of the equation.

However, if you specify the equation as

$$x+1/x = c(1) + c(2)*y$$

EViews will not be able to normalize the dependent variable and the Forecast dialog looks like this:



The dialog only allows you to forecast the transformed dependent variable, since EViews does not know how to normalize and solve for  $X$ . Note also that only static forecasts are available for this case. This restriction holds since EViews will not be able to solve for any lagged values of  $X$  on the right hand-side.

If the formula can be normalized, EViews will compute the forecasts of the transformed dependent variable by transforming the forecasts of the normalized series. This has important consequences when the formula includes lagged series. For example, consider the following two models:

```
series dy = d(y)
equation eq1.ls d(y) c x
equation eq2.ls dy c x
```

The dynamic forecasts of the first difference  $D(Y)$  from the first equation will be numerically identical to those for  $DY$  from the second equation. However, the static forecasts for  $D(Y)$  from the two equations will not be identical. This is because in the first equation, EViews knows that the dependent variable is a transformation of  $Y$ , so it will use the actual lagged value of  $Y$  in computing the static forecast of the first difference  $D(Y)$ . In the second equation, EViews simply views  $DY$  as an ordinary series, so that  $C$  and  $X$  are used to compute the static forecast.

### Plotted Standard Errors

When you select **Do graph** in the forecast dialog, EViews will plot the forecasts, along with plus and minus two standard error bands. When you estimate an equation with an expression for the left-hand side, EViews will plot the standard error bands for either the normalized or the unnormalized expression, depending upon which term you elect to forecast.

If you elect to predict the normalized dependent variable, EViews will automatically account for the nonlinearity in the standard error transformation. The next section provides additional details on the procedure used to normalize the upper and lower bounds.

### Saved Forecast Standard Errors

If you provide a name in this edit box, EViews will store the standard errors of the underlying series or expression that you chose to forecast.

When the dependent variable of the equation is a simple series or a formula involving only *linear* transformations, the saved standard errors will be exact (except where the forecasts do not account for coefficient uncertainty, as described below). If the dependent variable involves non-linear transformations, the saved forecast standard errors will be exact if you choose to forecast the entire formula. If you choose to forecast the underlying endogenous series, the forecast uncertainty cannot be computed exactly, and EViews will provide a linear (first-order) approximation to the forecast standard errors.

Consider the following equations involving a formula dependent variable:

$$d(y) \propto x$$

$$\log(y) \propto x$$

For the first equation you may choose to forecast either  $Y$  or  $D(Y)$ . In both cases, the forecast standard errors will be exact, since the expression involves only linear transformations. The two standard errors will, however, differ in dynamic forecasts since the forecast standard errors for  $Y$  take into account the forecast uncertainty from the lagged value of  $Y$ . In the second example, the forecast standard errors for  $\log(Y)$  will be exact. If, however, you request a forecast for  $Y$  itself, the standard errors saved in the series will be the approximate (linearized) forecast standard errors for  $Y$ .

Note that when EViews displays a graph view of the forecasts together with standard error bands, the standard error bands are always exact. Thus, in forecasting the underlying dependent variable in a nonlinear expression, the standard error bands will not be the same as those you would obtain by constructing series using the linearized standard errors saved in the workfile.

Suppose in our second example above that you store the forecast of  $Y$  and its standard errors in the workfile as the series `YHAT` and `SE_YHAT`. Then the *approximate* two standard error bounds can be generated manually as:

```
series yhat_high1 = yhat + 2*se_yhat
series yhat_low1 = yhat - 2*se_yhat
```

These forecast error bounds will be symmetric about the point forecasts `YHAT`.

On the other hand, when EViews plots the forecast error bounds of  $Y$ , it proceeds in two steps. It first obtains the forecast of  $\log(Y)$  and its standard errors (say `LYHAT` and `SE_LYHAT`) and forms the forecast error bounds on  $\log(Y)$ :

```
lyhat + 2*se_lyhat
lyhat - 2*se_lyhat
```

It then normalizes (inverts the transformation) of the two standard error bounds to obtain the prediction interval for  $Y$ :

```
series yhat_high2 = exp(lyhat + 2*se_lyhat)
series yhat_low2 = exp(lyhat - 2*se_lyhat)
```

Because this transformation is a non-linear transformation, these bands will not be symmetric around the forecast.

To take a more complicated example, suppose that you generate the series `DLY` and `LY`, and then estimate three equivalent models:



```
series dly =dlog(y)
series ly = log(y)
equation eq1.ls dlog(y) c x
equation eq2.ls d(ly) c x
equation eq3.ls dly c x
```

The estimated equations from the three models are numerically identical. If you choose to forecast the underlying dependent (normalized) series from each model, EQ1 will forecast Y, EQ2 will forecast LY (the log of Y), and EQ3 will forecast DLY (the log of the first difference of Y,  $\log(Y) - \log(Y-1)$ ). The forecast standard errors saved from EQ1 will be linearized approximations to the forecast standard error of Y, while those from the latter two will be exact for the forecast standard error of log Y and the log of the first difference of Y.

Static forecasts from all three models are identical because the forecasts from previous periods are not used in calculating this period's forecast when performing static forecasts. For dynamic forecasts, the log of the forecasts from EQ1 will be identical to those from EQ2 and the log first difference of the forecasts from EQ1 will be identical to the first difference of the forecasts from EQ2 and to the forecasts from EQ3. For static forecasts, the log first difference of the forecasts from EQ1 will be identical to the first difference of the forecasts from EQ2. However, these forecasts differ from those obtained from EQ3 because EViews does not know that the generated series DLY is actually a difference term so that it does not use the dynamic relation in the forecasts.

A final word of caution. When you have lagged dependent variables, you should avoid referring to the lagged series before the current series in a dependent variable expression. For example, consider the two equation specifications

$$\begin{aligned}d(y) & c x \\ -y(-1) + y & c x\end{aligned}$$

Both models have the first difference of Y as the dependent variable and the estimation results are identical for the two models. However, if you forecast Y from the second model, EViews will try to calculate the forecasts of Y using leads of the actual series Y. These forecasts of Y will differ from those produced by the first model, and may not be what you expected.

## Forecasting with Nonlinear and PDL Specifications

As explained above, forecast errors can arise from two sources: coefficient uncertainty and innovation uncertainty. For linear regression models, the forecast standard errors account for both coefficient and innovation uncertainty. However, if the model is nonlinear in the parameters (or if it contains a PDL specification), then the standard errors ignore coefficient uncertainty. EViews will display a message in the status line at the bottom of the EViews window when forecast standard errors only account for innovation uncertainty.

For example, consider the three specifications

```
log(y) c x
y = c(1) + c(2)*x
y = exp(c(1)*x)
y c x pdl(z, 4, 2)
```

Forecast standard errors from the first and second models account for both coefficient and innovation uncertainty since both models are linear in the coefficients. The third and fourth specifications have forecast standard errors that account only for residual uncertainty.

One additional case requires mention. Suppose you have the specification:

```
y-c(1) = c(3) + c(2)*x
```

Despite the fact that this specification is linear in the parameters, EViews will ignore coefficient uncertainty. Forecast standard errors for any specification that contains coefficients on the left-hand side of the equality will only reflect residual uncertainty.

## Commands

To obtain static (one-step ahead) forecasts, follow the name of the estimated equation, a dot, the command `fit`, a name for the fitted series, and optionally a name for the standard errors of the fitted values:

```
eq1.fit yhat yhat_se
```

To obtain dynamic forecasts, follow the name of the estimated equation, a dot, the command `forecast`, a name for the forecasts, and optionally a name for the standard errors of the forecasts:

```
eq1.forecast yh yh_se
```

See the *Command and Programming Reference* for a complete list of commands and options available for forecasting.

## Chapter 15. Specification and Diagnostic Tests

---

Empirical research is usually an interactive process. The process begins with a specification of the relationship to be estimated. Selecting a specification usually involves several choices: the variables to be included, the functional form connecting these variables, and if the data are time series, the dynamic structure of the relationship between the variables.

Inevitably, there is uncertainty regarding the appropriateness of this initial specification. Once you estimate your equation, EViews provides tools for evaluating the quality of your specification along a number of dimensions. In turn, the results of these tests influence the chosen specification, and the process is repeated.

This chapter describes the extensive menu of specification test statistics that are available as views or procedures of an equation object. While we attempt to provide you with sufficient statistical background to conduct the tests, practical considerations ensure that many of the descriptions are incomplete. We refer you to standard statistical and econometric references for further details.

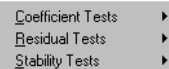
### Background

Each test procedure described below involves the specification of a null hypothesis, which is the hypothesis under test. Output from a test command consists of the sample values of one or more test statistics and their associated probability numbers ( $p$ -values). The latter indicate the probability of obtaining a test statistic whose absolute value is greater than or equal to that of the sample statistic if the null hypothesis is true. Thus, low  $p$ -values lead to the rejection of the null hypothesis. For example, if a  $p$ -value lies between 0.05 and 0.01, the null hypothesis is rejected at the 5 percent but not at the 1 percent level.

Bear in mind that there are different assumptions and distributional results associated with each test. For example, some of the test statistics have exact, finite sample distributions (usually  $t$  or  $F$ -distributions). Others are large sample test statistics with asymptotic  $\chi^2$  distributions. Details vary from one test to another and are given below in the description of each test.

### Types of Tests

The **View** button on the equation toolbar gives you a choice among three categories of tests to check the specification of the equation.



Additional tests are discussed elsewhere in the *User's Guide*. These include unit root tests (Chapter 13, page 298), the Granger causality test (Chapter 8, page 193), tests specific to binary, order, censored, and count models (Chapter 17), and the Johansen test for cointegration (Chapter 20, page 488).

## Coefficient Tests

Wald - Coefficient Restrictions...  
Omitted Variables - Likelihood Ratio...  
Redundant Variables - Likelihood Ratio...

These tests evaluate restrictions on the estimated coefficients, including the special case of tests for omitted and redundant variables.

### Wald Test — Coefficient Restrictions

The Wald test computes the test statistic by estimating the unrestricted regression without imposing the coefficient restrictions specified by the null hypothesis. The Wald statistic measures how close the unrestricted estimates come to satisfying the restrictions under the null hypothesis. If the restrictions are in fact true, then the unrestricted estimates should come close to satisfying the restrictions.

We first give a general formula for the computation of the Wald test statistic; users who are familiar with the formula or wish to skip the technical details can skip this subsection.

#### Details

Consider a general nonlinear regression model

$$y = f(\beta) + \epsilon \quad (15.1)$$

where  $y$  and  $\epsilon$  are  $T$ -vectors and  $\beta$  is a  $k$  vector of parameters to be estimated. Any restrictions on the parameters can be written as

$$H_0: g(\beta) = 0, \quad (15.2)$$

where  $g$  is a smooth function,  $g: \mathbb{R}^k \rightarrow \mathbb{R}^q$ , imposing  $q$  restrictions on  $\beta$ . The Wald statistic is then computed as

$$W = Tg(b)' \left( \frac{\partial g(b)}{\partial b} V(b) \frac{\partial g(b)}{\partial b'} \right) g(b) \quad (15.3)$$

where  $T$  is the number of observations and  $b$  is the vector of unrestricted parameter estimates.  $V$  is the estimated variance of  $b$  given by

$$V = Ts^2 \left( \frac{\partial g(b)}{\partial b} \frac{\partial g(b)}{\partial b'} \right)^{-1} \quad (15.4)$$

where  $u$  is the vector of unrestricted residuals, and  $s^2$  is the usual estimator of the unrestricted residual variance,  $s^2 = (u'u)/(N-k)$

More formally, under the null hypothesis  $H_0$ , the Wald statistic has an asymptotic  $\chi^2(q)$  distribution, where  $q$  is the number of restrictions under  $H_0$ .

The Wald test can be applied to equations estimated by least squares, two-stage least squares, nonlinear least squares, binary, ordered, censored, truncated, and count models, and systems estimators. The Wald test is the only test described in this chapter that can be applied to equations estimated by system methods.

For the special case of a linear regression model

$$y = X\beta + \epsilon \quad (15.5)$$

and linear restrictions

$$H_0: R\beta - r = 0, \quad (15.6)$$

where  $R$  is a known  $q \times k$  matrix, and  $r$  is a  $q$ -vector, respectively. The Wald statistic in Equation (15.3) reduces to

$$W = (Rb - r)'(s^2 R(X'X)^{-1} R)'(Rb - r), \quad (15.7)$$

which is asymptotically distributed as  $\chi^2(q)$  under  $H_0$ .

If we further assume that the errors  $\epsilon$  are independent and identically normally distributed, we have an exact, finite sample  $F$ -statistic:

$$F = \frac{(\tilde{u}'\tilde{u} - u'u)/q}{(u'u)/(T - k)} = \frac{W}{q}, \quad (15.8)$$

where  $\tilde{u}$  is the vector of residuals from the restricted regression. The  $F$ -statistic compares the residual sum of squares computed with and without the restrictions imposed. If the restrictions are valid, there should be little difference in the two residual sum of squares and the  $F$ -value should be small.

EViews reports both the chi-square and the  $F$ -statistics and the associated  $p$ -values.

### How to Perform Wald Coefficient Tests

To demonstrate how to perform Wald tests, we consider an example. Suppose a Cobb-Douglas production function has been estimated in the form:

$$\log Q = A + \alpha \log L + \beta \log K + \epsilon, \quad (15.9)$$

where  $Q$ ,  $K$  and  $L$  denote value-added output and the inputs of capital and labor respectively. The hypothesis of constant returns to scale is then tested by the restriction:  $\alpha + \beta = 1$ .

Estimation of the Cobb-Douglas production function using annual data from 1947 to 1971 provided the following result:

Dependent Variable: LOG(Q)  
 Method: Least Squares  
 Date: 08/11/97 Time: 16:56  
 Sample: 1947 1971  
 Included observations: 25

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-2.327939	0.410601	-5.669595	0.0000
LOG(L)	1.591175	0.167740	9.485970	0.0000
LOG(K)	0.239604	0.105390	2.273498	0.0331
R-squared	0.983672	Mean dependent var	4.767586	
Adjusted R-squared	0.982187	S.D. dependent var	0.326086	
S.E. of regression	0.043521	Akaike info criterion	-3.318997	
Sum squared resid	0.041669	Schwarz criterion	-3.172732	
Log likelihood	44.48746	F-statistic	662.6819	
Durbin-Watson stat	0.637300	Prob(F-statistic)	0.000000	

The sum of the coefficients on LOG(L) and LOG(K) appears to be in excess of one, but to determine whether the difference is statistically relevant, we will conduct the hypothesis test of constant returns.

To carry out a Wald test, choose **View/Coefficient Tests/Wald-Coefficient Restrictions...** from the equation toolbar. Enter the restrictions into the edit box, with multiple coefficient restrictions separated by commas. The restrictions should be expressed as equations involving the estimated coefficients and constants (you may not include series names). The coefficients should be referred to as C(1), C(2), and so on, unless you have used a different coefficient vector in estimation.

To test the hypothesis of constant returns to scale, type the following restriction in the dialog box:

$$c(2) + c(3) = 1$$

and click **OK**. EViews reports the following result of the Wald test:

Wald Test:			
Equation: EQ1			
Null Hypothesis: C(2)+C(3)=1			
F-statistic	120.0177	Probability	0.000000
Chi-square	120.0177	Probability	0.000000

EViews reports an  $F$ -statistic and a Chi-square statistic with associated  $p$ -values. The Chi-square statistic is equal to the  $F$ -statistic times the number of restrictions under test. In this example, there is only one restriction and so the two test statistics are identical with the  $p$ -values of both statistics indicating that we can decisively reject the null hypothesis of constant returns to scale.

To test more than one restriction, separate the restrictions by commas. For example, to test the hypothesis that the elasticity of output with respect to labor is  $2/3$  and with respect to capital is  $1/3$ , type the restrictions as

$$c(2)=2/3, \quad c(3)=1/3$$

and EViews reports

Wald Test:			
Equation: DEMAND			
Null Hypothesis: C(2)=2/3			
C(1)=1/3			
F-statistic	385.6769	Probability	0.000000
Chi-square	771.3538	Probability	0.000000

As an example of a nonlinear model with a nonlinear restriction, we estimate a production function of the form

$$\log Q = \beta_1 + \beta_2 \log(\beta_3 K^{\beta_4} + (1 - \beta_3) L^{\beta_4}) + \epsilon \quad (15.10)$$

and test the constant elasticity of substitution (CES) production function restriction  $\beta_2 = 1/\beta_4$ . This is an example of a nonlinear restriction. To estimate the (unrestricted) nonlinear model, you should select **Quick/Estimate Equation...** and then enter the following specification:

$$\log(q) = c(1) + c(2) * \log(c(3) * k^{c(4)} + (1 - c(3)) * l^{c(4)})$$

To test the nonlinear restriction, choose **View/Coefficient Tests/Wald-Coefficient Restrictions...** from the equation toolbar and type the following restriction in the Wald Test dialog box:

$$c(2)=1/c(4)$$

The results are presented below:

Wald Test:			
Equation: EQ2			
Null Hypothesis: C(2)=1/C(4)			
F-statistic	0.028507	Probability	0.867539
Chi-square	0.028507	Probability	0.865923

Since this is a nonlinear test, we focus on the Chi-square statistic which fails to reject the null hypothesis.

It is well-known that nonlinear Wald tests are not invariant to the way that you specify the nonlinear restrictions. In this example, the nonlinear restriction  $\beta_2 = 1/\beta_4$  can equivalently be written as  $\beta_2\beta_4 = 1$  or  $\beta_4 = 1/\beta_2$ . For example, entering the restriction as

$$c(2) * c(4) = 1$$

yields:

Wald Test:			
Equation: EQ2			
Null Hypothesis: C(2)*C(4)=1			
F-statistic	104.5599	Probability	0.000000
Chi-square	104.5599	Probability	0.000000

The test now decisively rejects the null hypothesis. We hasten to add that this is not a situation that is unique to EViews, but is a more general property of the Wald test. Unfortunately, there does not seem to be a general solution to this problem; see Davidson and MacKinnon (1993, Chapter 13) for further discussion and references.

An alternative is to use the LR test, which is invariant to reparameterization. To carry out the LR test, you compute both the unrestricted model (as estimated above) and the restricted model (imposing the nonlinear restrictions) and compare the log likelihood values. We will estimate the two specifications using the command form. The restricted model can be estimated by the entering the following command in the command window:

$$\text{equation eq\_ces0.ls log(q) = c(1) + c(2)*log(c(3)*k^{(1/c(2))} + (1 - c(3))*1^{(1/c(2))})}$$

while the unrestricted model is estimated using the command:

$$\text{equation eq\_ces1.ls log(q) = c(1) + c(2)*log(c(3)*k^{c(4)} + (1 - c(3))*1^{c(4)})}$$

Note that we save the two equations in the named equation objects EQ0 and EQ1. The LR test statistic can then be computed as

$$\text{scalar lr} = -2 * (\text{eq\_ces0.@logl} - \text{eq\_ces1.@logl})$$

To see the value of LR, double click on LR; the value will be displayed in the status line at the bottom of the EViews window. You can also compute the  $p$ -value of LR by the command

$$\text{scalar lr\_pval} = 1 - @cchisq(lr, 1)$$



## Omitted Variables

This test enables you to add a set of variables to an existing equation and to ask whether the set makes a significant contribution to explaining the variation in the dependent variable. The null hypothesis  $H_0$  is that the additional set of regressors are not jointly significant.

The output from the test is an  $F$ -statistic and a likelihood ratio (LR) statistic with associated  $p$ -values, together with the estimation results of the unrestricted model under the alternative. The  $F$ -statistic is based on the difference between the residual sums of squares of the restricted and unrestricted regressions. The LR statistic is computed as

$$LR = -2(l_r - l_u) \quad (15.11)$$

where  $l_r$  and  $l_u$  are the maximized values of the (Gaussian) log likelihood function of the unrestricted and restricted regressions, respectively. Under  $H_0$ , the LR statistic has an asymptotic  $\chi^2$  distribution with degrees of freedom equal to the number of restrictions (the number of added variables).

Bear in mind that:

- The omitted variables test requires that the same number of observations exist in the original and test equations. If any of the series to be added contain missing observations over the sample of the original equation (which will often be the case when you add lagged variables), the test statistics cannot be constructed.
- The omitted variables test can be applied to equations estimated with linear LS, TSLS, ARCH (mean equation only), binary, ordered, censored, truncated, and count models. The test is available only if you specify the equation by listing the regressors, not by a formula.

To perform an LR test in these settings, you can estimate a separate equation for the unrestricted and restricted models over a common sample, and evaluate the LR statistic and  $p$ -value using scalars and the `@cchisq` function, as described above.

### How to Perform an Omitted Variables Test

To test for omitted variables, select **View/Coefficient Tests/Omitted Variables-Likelihood Ratio...** In the dialog that opens, list the names of the test variables, each separated by at least one space. Suppose, for example, that the initial regression is

```
ls log(q) c log(l) log(k)
```

If you enter the list

```
log(m) log(e)
```

in the dialog, then EViews reports the results of the unrestricted regression containing the two additional explanatory variables, and displays statistics testing the hypothesis that the coefficients on the new variables are jointly zero. The top part of the output depicts the test results:

Omitted Variables: LOG(M) LOG(E)			
F-statistic	4.267478	Probability	0.028611
Log likelihood ratio	8.884940	Probability	0.011767

The  $F$ -statistic has an exact finite sample  $F$ -distribution under  $H_0$  if the errors are independent and identically distributed normal random variables. The numerator degrees of freedom is the number of additional regressors and the denominator degrees of freedom is the number of observations less the total number of regressors. The log likelihood ratio statistic is the LR test statistic and is asymptotically distributed as a  $\chi^2$  with degrees of freedom equal to the number of added regressors.

In our example, the tests reject the null hypothesis that the two series do not belong to the equation at a 5% significance level, but cannot reject the hypothesis at a 1% significance level.

## Redundant Variables

The redundant variables test allows you to test for the statistical significance of a subset of your included variables. More formally, the test is for whether a subset of variables in an equation all have zero coefficients and might thus be deleted from the equation. The redundant variables test can be applied to equations estimated by linear LS, TSLS, ARCH (mean equation only), binary, ordered, censored, truncated, and count methods. The test is available only if you specify the equation by listing the regressors, not by a formula.

### How to Perform a Redundant Variables Test

To test for redundant variables, select **View/Coefficient Tests/Redundant Variables-Likelihood Ratio...** In the dialog that appears, list the names of each of the test variables, separated by at least one space. Suppose, for example, that the initial regression is

$$ls \log(q) \ c \log(l) \ \log(k) \ \log(m) \ \log(e)$$

If you type the list

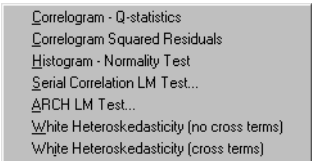
$$\log(m) \ \log(e)$$

in the dialog, then EViews reports the results of the restricted regression dropping the two regressors, followed by the statistics associated with the test of the hypothesis that the coefficients on the two variables are jointly zero.

The test statistics are the  $F$ -statistic and the Log likelihood ratio. The  $F$ -statistic has an exact finite sample  $F$ -distribution under  $H_0$  if the errors are independent and identically distributed normal random variables. The numerator degrees of freedom are given by the number of coefficient restrictions in the null hypothesis. The denominator degrees of freedom are given by the total regression degrees of freedom. The LR test is an asymptotic test, distributed as a  $\chi^2$  with degrees of freedom equal to the number of excluded variables under  $H_0$ . In this case, there are two degrees of freedom.

## Residual Tests

EViews provides tests for serial correlation, normality, heteroskedasticity, and autoregressive conditional heteroskedasticity in the residuals from your estimated equation. Not all of these tests are available for every specification.



Correlogram - Q-statistics  
Correlogram Squared Residuals  
Histogram - Normality Test  
Serial Correlation LM Test...  
ARCH LM Test...  
White Heteroskedasticity (no cross terms)  
White Heteroskedasticity (cross terms)

### Correlograms and Q-statistics

This view displays the autocorrelations and partial autocorrelations of the equation residuals up to the specified number of lags. Further details on these statistics and the Ljung-Box  $Q$ -statistics that are also computed are provided in Chapter 7, “ $Q$ -Statistics” on page 159.

This view is available for the residuals from least squares, two-stage least squares, nonlinear least squares and binary, ordered, censored, and count models. In calculating the probability values for the  $Q$ -statistics, the degrees of freedom are adjusted to account for estimated ARMA terms.

To display the correlograms and  $Q$ -statistics, push **View/Residual Tests/Correlogram-Q-statistics** on the equation toolbar. In the Lag Specification dialog box, specify the number of lags you wish to use in computing the correlogram.

### Correlograms of Squared Residuals

This view displays the autocorrelations and partial autocorrelations of the squared residuals up to any specified number of lags and computes the Ljung-Box  $Q$ -statistics for the corresponding lags. The correlograms of the squared residuals can be used to check autoregressive conditional heteroskedasticity (ARCH) in the residuals; see also “ARCH LM Test” on page 339, below.

If there is no ARCH in the residuals, the autocorrelations and partial autocorrelations should be zero at all lags and the  $Q$ -statistics should not be significant; see Chapter 7, page 156, for a discussion of the correlograms and  $Q$ -statistics.

This view is available for equations estimated by least squares, two-stage least squares, and nonlinear least squares estimation. In calculating the probability for  $Q$ -statistics, the degrees of freedom are adjusted for the inclusion of ARMA terms.

To display the correlograms and  $Q$ -statistics of the squared residuals, push **View/Residual Tests/Correlogram Squared Residuals** on the equation toolbar. In the Lag Specification dialog box that opens, specify the number of lags over which to compute the correlograms.

## Histogram and Normality Test

This view displays a histogram and descriptive statistics of the residuals, including the Jarque-Bera statistic for testing normality. If the residuals are normally distributed, the histogram should be bell-shaped and the Jarque-Bera statistic should not be significant; see Chapter 7, page 145, for a discussion of the Jarque-Bera test. This view is available for residuals from least squares, two-stage least squares, nonlinear least squares, and binary, ordered, censored, and count models.

To display the histogram and Jarque-Bera statistic, select **View/Residual Tests/Histogram-Normality**. The Jarque-Bera statistic has a  $\chi^2$  distribution with two degrees of freedom under the null hypothesis of normally distributed errors.

## Serial Correlation LM Test

This test is an alternative to the  $Q$ -statistics for testing serial correlation. The test belongs to the class of asymptotic (large sample) tests known as Lagrange multiplier (LM) tests.

Unlike the Durbin-Watson statistic for AR(1) errors, the LM test may be used to test for higher order ARMA errors, and is applicable whether or not there are lagged dependent variables. Therefore, we recommend its use whenever you are concerned with the possibility that your errors exhibit autocorrelation.

The null hypothesis of the LM test is that there is no serial correlation up to lag order  $p$ , where  $p$  is a pre-specified integer. The local alternative is ARMA( $r, q$ ) errors, where the number of lag terms  $p = \max(r, q)$ . Note that the alternative includes both AR( $p$ ) and MA( $p$ ) error processes, and that the test may have power against a variety of autocorrelation structures. See Godfrey, 1988, for a discussion.

The test statistic is computed by an auxiliary regression as follows. First, suppose you have estimated the regression

$$y_t = X_t\beta + \epsilon_t \quad (15.12)$$

where  $b$  are the estimated coefficients and  $e$  are the residuals. The test statistic for lag order  $p$  is based on the auxiliary regression

$$e_t = X_t\gamma + \left( \sum_{s=1}^p \alpha_s e_{t-s} \right) + v_t \quad (15.13)$$

This is a regression of the residuals on the original regressors  $X$  and lagged residuals up to order  $p$ . EViews reports two test statistics from this test regression. The  $F$ -statistic is an omitted vari-

able test for the joint significance of all lagged residuals. Because the omitted variables are residuals and not independent variables, the exact finite sample distribution of the  $F$ -statistic under  $H_0$  is not known, but we still present the  $F$ -statistic for comparison purposes.

The Obs\*R-squared statistic is the Breusch-Godfrey LM test statistic. This LM statistic is computed as the number of observations, times the (uncentered)  $R^2$  from the test regression. Under quite general conditions, the LM test statistic is asymptotically distributed as a  $\chi^2(p)$ .

The serial correlation LM test is available for residuals from least squares or two-stage least squares. The original regression may include AR and MA terms, in which case the test regression will be modified to take account of the ARMA terms.

To carry out the test, push **View/Residual Tests/Serial Correlation LM Test...** on the equation toolbar and specify the highest order of the AR or MA process that might describe the serial correlation. If the test indicates serial correlation in the residuals, LS standard errors are invalid and should not be used for inference.

### ARCH LM Test

This is a Lagrange multiplier (LM) test for autoregressive conditional heteroskedasticity (ARCH) in the residuals (Engle 1982). This particular specification of heteroskedasticity was motivated by the observation that in many financial time series, the magnitude of residuals appeared to be related to the magnitude of recent residuals. ARCH in itself does not invalidate standard LS inference. However, ignoring ARCH effects may result in loss of efficiency; see Chapter 16 for a discussion of estimation of ARCH models in EViews.

The ARCH LM test statistic is computed from an auxiliary test regression. To test the null hypothesis that there is no ARCH up to order  $q$  in the residuals, we run the regression

$$e_t^2 = \beta_0 + \left( \sum_{s=1}^q \beta_s e_{t-s}^2 \right) + v_t, \quad (15.14)$$

where  $e$  is the residual. This is a regression of the squared residuals on a constant and lagged squared residuals up to order  $q$ . EViews reports two test statistics from this test regression. The  $F$ -statistic is an omitted variable test for the joint significance of all lagged squared residuals. The Obs\*R-squared statistic is Engle's LM test statistic, computed as the number of observations times the  $R^2$  from the test regression. The exact finite sample distribution of the  $F$ -statistic under  $H_0$  is not known but the LM test statistic is asymptotically distributed  $\chi^2(q)$  under quite general conditions. The ARCH LM test is available for equations estimated by least squares, two-stage least squares, and nonlinear least squares.

To carry out the test, push **View/Residual Tests/ARCH LM Test...** on the equation toolbar and specify the order of ARCH to be tested against.

## White's Heteroskedasticity Test

This is a test for heteroskedasticity in the residuals from a least squares regression (White, 1980). Ordinary least squares estimates are consistent in the presence heteroskedasticity, but the conventional computed standard errors are no longer valid. If you find evidence of heteroskedasticity, you should either choose the robust standard errors option to correct the standard errors (see Chapter 12, page 251) or you should model the heteroskedasticity to obtain more efficient estimates using weighted least squares.

White's test is a test of the null hypothesis of no heteroskedasticity against heteroskedasticity of some unknown general form. The test statistic is computed by an auxiliary regression, where we regress the squared residuals on all possible (nonredundant) cross products of the regressors. For example, suppose we estimated the following regression:

$$y_t = b_1 + b_2x_t + b_3z_t + e_t. \quad (15.15)$$

where the  $b$  are the estimated parameters and  $e$  the residual. The test statistic is then based on the auxiliary regression:

$$e_t^2 = \alpha_0 + \alpha_1x_t + \alpha_2z_t + \alpha_3x_t^2 + \alpha_4z_t^2 + \alpha_5x_tz_t + v_t. \quad (15.16)$$

EViews reports two test statistics from the test regression. The  $F$ -statistic is an omitted variable test for the joint significance of all cross products, excluding the constant. It is presented for comparison purposes.

The Obs\*R-squared statistic is White's test statistic, computed as the number of observations times the centered  $R^2$  from the test regression. The exact finite sample distribution of the  $F$ -statistic under  $H_0$  is not known, but White's test statistic is asymptotically distributed as a  $\chi^2$  with degrees of freedom equal to the number of slope coefficients (excluding the constant) in the test regression.

White also describes this approach as a general test for model misspecification, since the null hypothesis underlying the test assumes that the errors are both homoskedastic and independent of the regressors, and that the linear specification of the model is correct. Failure of any one of these conditions could lead to a significant test statistic. Conversely, a non-significant test statistic implies that none of the three conditions is violated.

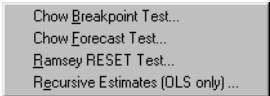
When there are redundant cross-products, EViews automatically drops them from the test regression. For example, the square of a dummy variable is the dummy variable itself, so that EViews drops the squared term to avoid perfect collinearity.

To carry out White's heteroskedasticity test, select **View/Residual Tests/White Heteroskedasticity**. EViews has two options for the test: cross terms and no cross terms. The cross terms version of the test is the original version of White's test that includes all of the cross product terms (in the example above,  $x_tz_t$ ). However, with many right-hand side variables in the regression, the number of possible cross product terms becomes very large so that it may not be practical to

include all of them. The no cross terms option runs the test regression using only squares of the regressors.

## Specification and Stability Tests

EViews provides a number of test statistic views that examine whether the parameters of your model are stable across various subsamples of your data.



Chow Breakpoint Test...  
Chow Forecast Test...  
Ramsey RESET Test...  
Recursive Estimates (OLS only) ...

One recommended empirical technique is to split the  $T$  observations in your data set of observations into  $T_1$  observations to be used for estimation, and  $T_2 = T - T_1$  observations to be used for testing and evaluation. Using all available sample observations for estimation promotes a search for a specification that best fits that specific data set, but does not allow for testing predictions of the model against data that have not been used in estimating the model. Nor does it allow one to test for parameter constancy, stability and robustness of the estimated relationship. In time series work you will usually take the first  $T_1$  observations for estimation and the last  $T_2$  for testing. With cross section data you may wish to order the data by some variable, such as household income, sales of a firm, or other indicator variables and use a sub-set for testing.

There are no hard and fast rules for determining the relative sizes of  $T_1$  and  $T_2$ . In some cases there may be obvious points at which a break in structure might have taken place—a war, a piece of legislation, a switch from fixed to floating exchange rates, or an oil shock. Where there is no reason *a priori* to expect a structural break, a commonly used rule-of-thumb is to use 85 to 90 percent of the observations for estimation and the remainder for testing.

EViews provides built-in procedures which facilitate variations on this type of analysis.

### Chow's Breakpoint Test

The idea of the breakpoint Chow test is to fit the equation separately for each subsample and to see whether there are significant differences in the estimated equations. A significant difference indicates a structural change in the relationship. For example, you can use this test to examine whether the demand function for energy was the same before and after the oil shock. The test may be used with least squares and two-stage least squares regressions.

To carry out the test, we partition the data into two or more subsamples. Each subsample must contain more observations than the number of coefficients in the equation so that the equation can be estimated using each subsample. The Chow breakpoint test is based on a comparison of the sum of squared residuals obtained by fitting a single equation to the entire sample with the sum of squared residuals obtained when separate equations are fit to each subsample of the data.

EViews reports two test statistics for the Chow breakpoint test. The  $F$ -statistic is based on the comparison of the restricted and unrestricted sum of squared residuals and in the simplest case involving a single breakpoint, is computed as

$$F = \frac{(\tilde{u}'\tilde{u} - (u_1'u_1 + u_2'u_2))/k}{(u_1'u_1 + u_2'u_2)/(T-2k)}, \quad (15.17)$$

where  $\tilde{u}'\tilde{u}$  is the restricted sum of squared residuals,  $u_i'u_i$  is the sum of squared residuals from subsample  $i$ ,  $T$  is the total number of observations, and  $k$  is the number of parameters in the equation. This formula can be generalized naturally to more than one breakpoint. The  $F$ -statistic has an exact finite sample  $F$ -distribution if the errors are independent and identically distributed normal random variables.

The log likelihood ratio statistic is based on the comparison of the restricted and unrestricted maximum of the (Gaussian) log likelihood function. The LR test statistic has an asymptotic  $\chi^2$  distribution with degrees of freedom equal to  $(m-1)k$  under the null hypothesis of no structural change, where  $m$  is the number of subsamples.

One major drawback of the breakpoint test is that each subsample requires at least as many observations as the number of estimated parameters. This may be a problem if, for example, you want to test for structural change between wartime and peacetime where there are only a few observations in the wartime sample. The Chow forecast test, discussed below, should be used in such cases.

To apply the Chow breakpoint test, push **View/Stability Tests/Chow Breakpoint Test...** on the equation toolbar. In the dialog that appears, list the dates or observation numbers for the breakpoints. For example, if your original equation was estimated from 1950 to 1994, entering

1960

in the dialog specifies two subsamples, one from 1950 to 1959 and one from 1960 to 1994. Typing

1960 1970

specifies three subsamples, 1950 to 1959, 1960 to 1969, and 1970 to 1994.

### Chow's Forecast Test

The Chow forecast test estimate sthe model for a subsample comprised of the first  $T_1$  observations. The estimated model is then used to predict the values of the dependent variable in the remaining  $T_2$  data points. A large difference between the actual and predicted values casts doubt on the stability of the estimated relation over the two subsamples. The Chow forecast test can be used with least squares and two-stage least squares regressions.

EViews reports two test statistics for the Chow forecast test. The  $F$ -statistic is computed as



$$F = \frac{(\tilde{u}'\tilde{u} - u'u)/T_2}{u'u/(T_1 - k)}, \quad (15.18)$$

where  $\tilde{u}'\tilde{u}$  is the residual sum of squares when the equation is fitted to all  $T$  sample observations,  $u'u$  is the residual sum of squares when the equation is fitted to  $T_1$  observations, and  $k$  is the number of estimated coefficients. This  $F$ -statistic has an exact finite sample  $F$ -distribution only if the errors are independent, and identically, normally distributed.

The log likelihood ratio statistic is based on the comparison of the restricted and unrestricted maximum of the (Gaussian) log likelihood function. Both the restricted and unrestricted log likelihood are obtained by estimating the regression using the whole sample. The restricted regression uses the original set of regressors, while the unrestricted regression adds a dummy variable for each forecast point. The LR test statistic has an asymptotic  $\chi^2$  distribution with degrees of freedom equal to the number of forecast points  $T_2$  under the null hypothesis of no structural change.

To apply Chow's forecast test, push **View/Stability Tests/Chow Forecast Test...** on the equation toolbar and specify the date or observation number for the beginning of the forecasting sample. The date should be within the current sample of observations.

As an example, suppose we estimate a consumption function using quarterly data from 1947:1 to 1994:4 and specify 1973:1 as the first observation in the forecast period. The test reestimates the equation for the period 1947:1 to 1972:4, and uses the result to compute the prediction errors for the remaining quarters, and reports the following results:

Chow Forecast Test: Forecast from 1973:1 to 1994:4			
F-statistic	0.708348	Probability	0.951073
Log likelihood ratio	91.57088	Probability	0.376108

Neither of the forecast test statistics reject the null hypothesis of no structural change in the consumption function before and after 1973:1.

If we test the same hypothesis using the Chow breakpoint test, the result is

Chow Breakpoint Test: 1973:1			
F-statistic	38.39198	Probability	0.000000
Log likelihood ratio	65.75468	Probability	0.000000

Note that both of the breakpoint test statistics decisively reject the hypothesis from above. This example illustrates the possibility that the two Chow tests may yield conflicting results.

### Ramsey's RESET Test

RESET stands for *Regression Specification Error Test* and was proposed by Ramsey (1969). The classical normal linear regression model is specified as

$$y = X\beta + \epsilon, \quad (15.19)$$

where the disturbance vector  $\epsilon$  is presumed to follow the multivariate normal distribution  $N(0, \sigma^2 I)$ . Specification error is an omnibus term which covers any departure from the assumptions of the maintained model. Serial correlation, heteroskedasticity, or non-normality of all violate the assumption that the disturbances are distributed  $N(0, \sigma^2 I)$ . Tests for these specification errors have been described above. In contrast, RESET is a general test for the following types of specification errors:

- Omitted variables;  $X$  does not include all relevant variables.
- Incorrect functional form; some or all of the variables in  $y$  and  $X$  should be transformed to logs, powers, reciprocals, or in some other way.
- Correlation between  $X$  and  $\epsilon$ , which may be caused, among other things, by measurement error in  $X$ , simultaneity, or the presence of lagged  $y$  values and serially correlated disturbances.

Under such specification errors, LS estimators will be biased and inconsistent, and conventional inference procedures will be invalidated. Ramsey (1969) showed that any or all of these specification errors produce a non-zero mean vector for  $\epsilon$ . Therefore, the null and alternative hypotheses of the RESET test are

$$\begin{aligned} H_0: \epsilon &\sim N(0, \sigma^2 I) \\ H_1: \epsilon &\sim N(\mu, \sigma^2 I) \quad \mu \neq 0 \end{aligned} \quad (15.20)$$

The test is based on an augmented regression

$$y = X\beta + Z\gamma + \epsilon. \quad (15.21)$$

The test of specification error evaluates the restriction  $\gamma = 0$ . The crucial question in constructing the test is to determine what variables should enter the  $Z$  matrix. Note that the  $Z$  matrix may, for example, be comprised of variables that are not in the original specification, so that the test of  $\gamma = 0$  is simply the omitted variables test described above.

In testing for incorrect functional form, the nonlinear part of the regression model may be some function of the regressors included in  $X$ . For example, if a linear relation

$$y = \beta_0 + \beta_1 X + \epsilon, \quad (15.22)$$

is specified instead of the true relation

$$y = \beta_0 + \beta_1 X + \beta_2 X^2 + \epsilon \quad (15.23)$$

the augmented model has  $Z = X^2$  and we are back to the omitted variable case. A more general example might be the specification of an additive relation

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon \quad (15.24)$$

instead of the (true) multiplicative relation

$$y = \beta_0 X_1^{\beta_1} X_2^{\beta_2} + \epsilon \quad (15.25)$$

A Taylor series approximation of the multiplicative relation would yield an expression involving powers and cross-products of the explanatory variables. Ramsey's suggestion is to include powers of the predicted values of the dependent variable (which are, of course, linear combinations of powers and cross-product terms of the explanatory variables) in  $Z$ . Specifically, Ramsey suggests

$$Z = [\hat{y}^2, \hat{y}^3, \hat{y}^4, \dots] \quad (15.26)$$

where  $\hat{y}$  is the vector of fitted values from the regression of  $y$  on  $X$ . The superscripts indicate the powers to which these predictions are raised. The first power is not included since it is perfectly collinear with the  $X$  matrix.

Output from the test reports the test regression and the  $F$ -statistic and Log likelihood ratio for testing the hypothesis that the coefficients on the powers of fitted values are all zero. A study by Ramsey and Alexander (1984) showed that the RESET test could detect specification error in an equation which was known *a priori* to be misspecified but which nonetheless gave satisfactory values for all the more traditional test criteria—goodness of fit, test for first order serial correlation, high  $t$ -ratios.

To apply the test, select **View/Stability Tests/Ramsey RESET Test...** and specify the number of fitted terms to include in the test regression. The fitted terms are the powers of the fitted values from the original regression, starting with the square or second power. For example, if you specify 1, then the test will add  $\hat{y}^2$  in the regression and if you specify 2, then the test will add  $\hat{y}^2$  and  $\hat{y}^3$  in the regression and so on. If you specify a large number of fitted terms, EViews may report a near singular matrix error message since the powers of the fitted values are likely to be highly collinear. The Ramsey RESET test is applicable only to an equation estimated by least squares.

## Recursive Least Squares

In recursive least squares the equation is estimated repeatedly, using ever larger subsets of the sample data. If there are  $k$  coefficients to be estimated in the  $b$  vector, then the first  $k$  observations are used to form the first estimate of  $b$ . The next observation is then added to the data set and  $k + 1$  observations are used to compute the second estimate of  $b$ . This process is repeated until all the  $T$  sample points have been used, yielding  $T - k + 1$  estimates of the  $b$  vector. At

each step the last estimate of  $b$  can be used to predict the next value of the dependent variable. The one-step ahead forecast error resulting from this prediction, suitably scaled, is defined to be a *recursive residual*.

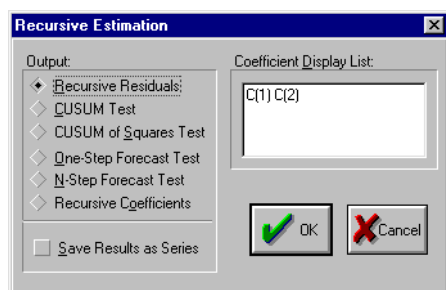
More formally, let  $X_{t-1}$  denote the  $t-1$  by  $k$  matrix of the regressors from period 1 to period  $t-1$ , and  $y_{t-1}$  the corresponding vector of observations on the dependent variable. These data up to period  $t-1$  give an estimated coefficient vector, denoted by  $b_{t-1}$ . This coefficient vector gives you a forecast of the dependent variable in period  $t$ . The forecast is  $x_t' b_{t-1}$ , where  $x_t'$  is the row vector of observations on the regressors in period  $t$ . The forecast error is  $y_t - x_t' b_{t-1}$ , and the forecast variance is given by:

$$\sigma^2 (1 + x_t' (X_{t-1}' X_{t-1})^{-1} x_t). \quad (15.27)$$

The recursive residual  $w_t$  is defined in EViews as

$$w_t = \frac{(y_t - x_t' b_{t-1})}{(1 + x_t' (X_{t-1}' X_{t-1})^{-1} x_t)^{1/2}} \quad (15.28)$$

These residuals can be computed for  $t = k+1, \dots, T$ . If the maintained model is valid, the recursive residuals will be independently and normally distributed with zero mean and constant variance  $\sigma^2$ .



To calculate the recursive residuals, press **View/ Stability Tests/Recursive Estimates (OLS only)**... on the equation toolbar. There are six options available for the recursive estimates view. The recursive estimates view is only available for equations estimated by ordinary least squares without AR and MA terms. The **Save Results as Series** option allows you to save the recursive residuals and recursive coefficients as named series in the workfile; see “Save Results as Series”

on page 349 below.

### Recursive Residuals

This option shows a plot of the recursive residuals about the zero line. Plus and minus two standard errors are also shown at each point. Residuals outside the standard error bands suggest instability in the parameters of the equation.

### CUSUM Test

The CUSUM test (Brown, Durbin, and Evans, 1975) is based on the cumulative sum of the recursive residuals. This option plots the cumulative sum together with the 5% critical lines. The test

finds parameter instability if the cumulative sum goes outside the area between the two critical lines.

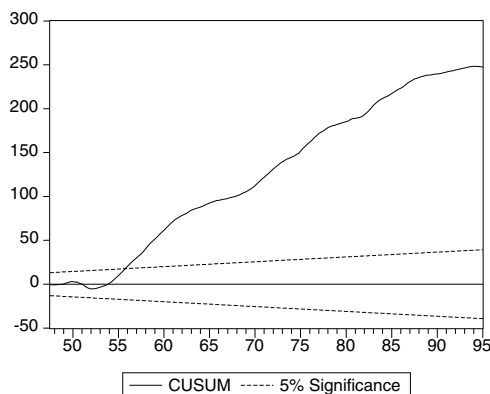
The CUSUM test is based on the statistic

$$W_t = \sum_{r=k+1}^t w_r / s, \quad (15.29)$$

for  $t = k + 1, \dots, T$ , where  $w$  is the recursive residual defined above, and  $s$  is the standard error of the regression fitted to all  $T$  sample points. If the  $\beta$  vector remains constant from period to period,  $E(W_t) = 0$ , but if  $\beta$  changes,  $W_t$  will tend to diverge from the zero mean value line. The significance of any departure from the zero line is assessed by reference to a pair of 5% significance lines, the distance between which increases with  $t$ . The 5% significance lines are found by connecting the points

$$[k, \pm 0.948(T-k)^{1/2}] \quad \text{and} \quad [T, \pm 3 \times 0.948(T-k)^{1/2}] \quad (15.30)$$

Movement of  $W_t$  outside the critical lines is suggestive of coefficient instability. A sample CUSUM test is given below.



The test clearly indicates instability in the equation during the sample period.

### CUSUM of Squares Test

The CUSUM of squares test (Brown, Durbin, and Evans, 1975) is based on the test statistic

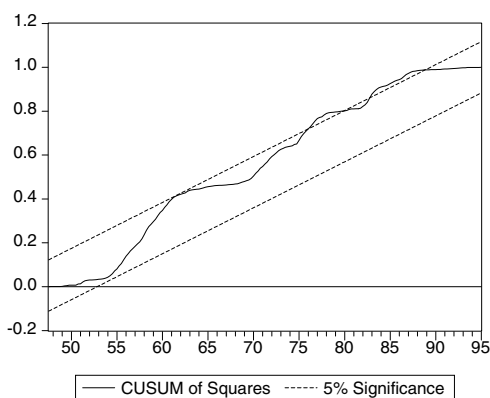
$$W_t = \left( \sum_{r=k+1}^t w_r^2 \right) / \left( \sum_{r=k+1}^T w_r^2 \right). \quad (15.31)$$

The expected value of  $S$  under the hypothesis of parameter constancy is

$$E(S_t) = (t-k)/(T-k), \quad (15.32)$$

which goes from zero at  $t = k$  to unity at  $t = T$ . The significance of the departure of  $S$  from its expected value is assessed by reference to a pair of parallel straight lines around the expected value. See Brown, Durbin, and Evans (1975) or Johnston and DiNardo (1997, Table D.8) for a table of significance lines for the CUSUM of squares test.

The CUSUM of squares test provides a plot of  $S_t$  against  $t$  and the pair of 5 percent critical lines. As with the CUSUM test, movement outside the critical lines is suggestive of parameter or variance instability.

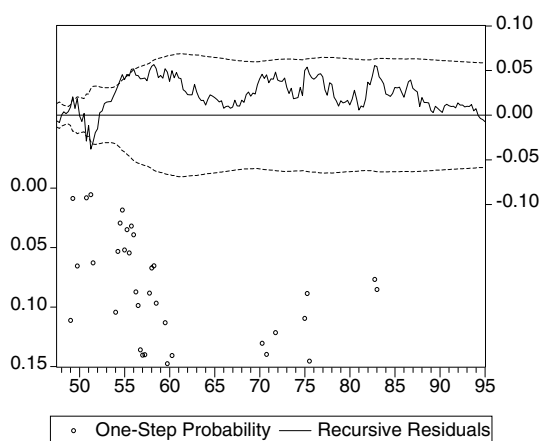


The cumulative sum of squares generally within the 5% significance lines, suggesting that the residual variance is somewhat stable.

### One-Step Forecast Test

If you look back at the definition of the recursive residuals given above, you will see that each recursive residual is the error in a one-step ahead forecast. To test whether the value of the dependent variable at time  $t$  might have come from the model fitted to all the data up to that point, each error can be compared with its standard deviation from the full sample.

The **One-Step Forecast Test** option produces a plot of the recursive residuals and standard errors and the sample points whose probability value is at or below 15 percent. The plot can help you spot the periods when your equation is least successful. For example, the one-step forecast test might look like this:



The upper portion of the plot (right vertical axis) repeats the recursive residuals and standard errors displayed by the **Recursive Residuals** option. The lower portion of the plot (left vertical axis) shows the probability values for those sample points where the hypothesis of parameter constancy would be rejected at the 5, 10, or 15 percent levels. The points with  $p$ -values less the 0.05 correspond to those points where the recursive residuals go outside the two standard error bounds.

For the test equation, there is evidence of instability early in the sample period.

### N-Step Forecast Test

This test uses the recursive calculations to carry out a sequence of Chow Forecast tests. In contrast to the single Chow Forecast test described earlier, this test does not require the specification of a forecast period— it automatically computes all feasible cases, starting with the smallest possible sample size for estimating the forecasting equation and then adding one observation at a time. The plot from this test shows the recursive residuals at the top and significant probabilities in the lower portion of the diagram.

### Recursive Coefficient Estimates

This view enables you to trace the evolution of estimates for any coefficient as more and more of the sample data are used in the estimation. The view will provide a plot of selected coefficients in the equation for all feasible recursive estimations. Also shown are the two standard error bands around the estimated coefficients.

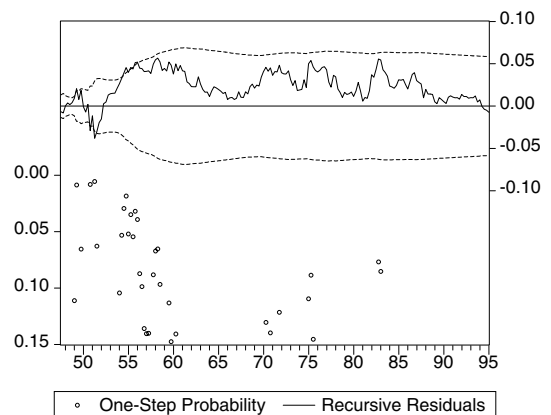
If the coefficient displays significant variation as more data is added to the estimating equation, it is a strong indication of instability. Coefficient plots will sometimes show dramatic jumps as the postulated equation tries to digest a structural break.

To view the recursive coefficient estimates, click the **Recursive Coefficients** option and list the coefficients you want to plot in the **Coefficient Display List** field of the dialog box. The recursive estimates of the marginal propensity to consume (coefficient C(2)), from the sample consumption function are provided below:

The estimated propensity to consume rises steadily as we add more data over the sample period, approaching a value of one.

### Save Results as Series

The **Save Results as Series** checkbox will do different things depending on the plot you have asked to be displayed. When paired with the **Recursive Coefficients** option, **Save Results as Series** will instruct EViews to save all recursive coefficients and their standard errors in the workfile as named series. EViews will name the coefficients using the next available name of the form, R\_C1, R\_C2, ..., and the corresponding standard errors as R\_C1SE, R\_C2SE, and so on.



If you check the **Save Results as Series** box with any of the other options, EViews saves the recursive residuals and the recursive standard errors as named series in the workfile. EViews will name the residual and standard errors as R\_RES and R\_RESSE, respectively.

Note that you can use the recursive residuals to reconstruct the CUSUM and CUSUM of squares series.

## Applications

In this section, we show how to carry out other specification tests in EViews. For brevity, the discussion is based on commands, but most of these procedures can also be carried out using the menu system.

### A Wald test of structural change with unequal variance

The  $F$ -statistics reported in the Chow tests have an  $F$ -distribution only if the errors are independent and identically normally distributed. This restriction implies that the residual variance in the two subsamples must be equal.

Suppose now that we wish to compute a Wald statistic for structural change with unequal subsample variances. Denote the parameter estimates and their covariance matrix in subsample  $i$  as  $b_i$  and  $V_i$  for  $i = 1, 2$ . Under the assumption that  $b_1$  and  $b_2$  are independent normal, the difference  $b_1 - b_2$  has mean zero and variance  $V_1 + V_2$ . Therefore a Wald statistic for the null hypothesis of no structural change and independent samples can be constructed as

$$W = (b_1 - b_2)'(V_1 + V_2)^{-1}(b_1 - b_2), \quad (15.33)$$

which has an asymptotic  $\chi^2$  distribution with degrees of freedom equal to the number of estimated parameters in the  $b$  vector.

To carry out this test in EViews, we estimate the model in each subsample and save the estimated coefficients and their covariance matrix. For example, suppose we have a quarterly sample of 1947:1–1994:4 and wish to test whether there was a structural change in the consumption function in 1973:1. First estimate the model in the first sample and save the results by the commands

```
coef(2) b1
smpl 1947:1 1972:4
equation eq_1.ls log(cs)=b1(1)+b1(2)*log(gdp)
sym v1=eq_1.@cov
```

The first line declares the coefficient vector, B1, into which we will place the coefficient estimates in the first sample. Note that the equation specification in the third line explicitly refers to elements of this coefficient vector. The last line saves the coefficient covariance matrix as a sym-



metric matrix named V1. Similarly, estimate the model in the second sample and save the results by the commands

```
coef(2) b2
smp1 1973.1 1994.4
equation eq_2.ls log(cs)=b2(1)+b2(2)*log(gdp)
sym v2=eq_2.@cov
```

To compute the Wald statistic, use the command

```
matrix wald=@transpose(b1-b2)*@inverse(v1+v2)*(b1-b2)
```

The Wald statistic is saved in the  $1 \times 1$  matrix named WALD. To see the value, either double click on WALD or type `show wald`. You can compare this value with the critical values from the  $\chi^2$  distribution with 2 degrees of freedom. Alternatively, you can compute the  $p$ -value in EViews using the command

```
scalar wald_p=1-@cchisq(wald(1,1),2)
```

The  $p$ -value is saved as a scalar named WALD\_P. To see the  $p$ -value, double click on WALD\_P or type `show wald_p`. The  $p$ -value will be displayed in the status line at the bottom of the EViews window.

## The Hausman test

A widely used class of tests in econometrics is the Hausman test. The underlying idea of the Hausman test is to compare two sets of estimates, one of which is consistent under both the null and the alternative and another which is consistent only under the null hypothesis. A large difference between the two sets of estimates is taken as evidence in favor of the alternative hypothesis.

Hausman (1978) originally proposed a test statistic for endogeneity based upon a direct comparison of coefficient values. Here we illustrate the version of the Hausman test proposed by Davidson and MacKinnon (1989, 1993), which carries out the test by running an auxiliary regression.

The following equation was estimated by OLS:

Dependent Variable: LOG(M1)  
Method: Least Squares  
Date: 08/13/97 Time: 14:12  
Sample(adjusted): 1959:02 1995:04  
Included observations: 435 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.022699	0.004443	-5.108528	0.0000
LOG(IP)	0.011630	0.002585	4.499708	0.0000
DLOG(PPI)	-0.024886	0.042754	-0.582071	0.5608
TB3	-0.000366	9.91E-05	-3.692675	0.0003
LOG(M1(-1))	0.996578	0.001210	823.4440	0.0000
R-squared	0.999953	Mean dependent var	5.844581	
Adjusted R-squared	0.999953	S.D. dependent var	0.670596	
S.E. of regression	0.004601	Akaike info criterion	-7.913714	
Sum squared resid	0.009102	Schwarz criterion	-7.866871	
Log likelihood	1726.233	F-statistic	2304897.	
Durbin-Watson stat	1.265920	Prob(F-statistic)	0.000000	

Suppose we are concerned that industrial production (IP) is endogenously determined with money (M1) through the money supply function. If this were the case, then OLS estimates will be biased and inconsistent. To test this hypothesis, we need to find a set of instrumental variables that are correlated with the “suspect” variable IP but not with the error term of the money demand equation. The choice of the appropriate instrument is a crucial step. Here we take the unemployment rate (URATE) and Moody’s AAA corporate bond yield (AAA) as instruments.

To carry out the Hausman test by artificial regression, we run two OLS regressions. In the first regression, we regress the suspect variable (log) IP on all exogenous variables and instruments and retrieve the residuals:

```
ls log(ip) c dlog(ppi) tb3 log(m1(-1)) urate aaa
series res_ip=resid
```

Then in the second regression, we re-estimate the money demand function including the residuals from the first regression as additional regressors. The result is:

Dependent Variable: LOG(M1)  
 Method: Least Squares  
 Date: 08/13/97 Time: 15:28  
 Sample(adjusted): 1959:02 1995:04  
 Included observations: 435 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.007145	0.007473	-0.956158	0.3395
LOG(IP)	0.001560	0.004672	0.333832	0.7387
DLOG(PPI)	0.020233	0.045935	0.440465	0.6598
TB3	-0.000185	0.000121	-1.527775	0.1273
LOG(M1(-1))	1.001093	0.002123	471.4894	0.0000
RES IP	0.014428	0.005593	2.579826	0.0102
R-squared	0.999954	Mean dependent var	5.844581	
Adjusted R-squared	0.999954	S.D. dependent var	0.670596	
S.E. of regression	0.004571	Akaike info criterion	-7.924512	
Sum squared resid	0.008963	Schwarz criterion	-7.868300	
Log likelihood	1729.581	F-statistic	1868171.	
Durbin-Watson stat	1.307838	Prob(F-statistic)	0.000000	

If the OLS estimates are consistent, then the coefficient on the first stage residuals should not be significantly different from zero. In this example, the test (marginally) rejects the hypothesis of consistent OLS estimates. (To be more precise, this is an asymptotic test and you should compare the  $t$ -statistic with the critical values from the standard normal.)

## Non-nested Tests

Most of the tests discussed in this chapter are nested tests in which the null hypothesis is obtained as a special case of the alternative hypothesis. Now consider the problem of choosing between the following two specifications of a consumption function:

$$\begin{aligned} H_1: \quad CS_t &= \alpha_1 + \alpha_2 GDP_t + \alpha_3 GDP_{t-1} + \epsilon_t \\ H_2: \quad CS_t &= \beta_1 + \beta_2 GDP_t + \beta_3 CS_{t-1} + \epsilon_t \end{aligned} \quad (15.34)$$

These are examples of non-nested models since neither model may be expressed as a restricted version of the other.

The  $J$ -test proposed by Davidson and MacKinnon (1993) provides one method of choosing between two non-nested models. The idea is that if one model is the correct model, then the fitted values from the other model should not have explanatory power when estimating that model. For example, to test model  $H_1$  against model  $H_2$ , we first estimate model  $H_2$  and retrieve the fitted values:

```
equation eq_cs2.ls cs c gdp cs(-1)
eq_cs2.fit cs2
```

The second line saves the fitted values as a series named CS2. Then estimate model  $H_1$  including the fitted values from model  $H_2$ . The result is:

Dependent Variable: CS  
Method: Least Squares  
Date: 8/13/97 Time: 00:49  
Sample(adjusted): 1947:2 1994:4  
Included observations: 191 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	7.313232	4.391305	1.665389	0.0975
GDP	0.278749	0.029278	9.520694	0.0000
GDP(-1)	-0.314540	0.029287	-10.73978	0.0000
CS2	1.048470	0.019684	53.26506	0.0000
R-squared	0.999833	Mean dependent var		1953.966
Adjusted R-squared	0.999830	S.D. dependent var		848.4387
S.E. of regression	11.05357	Akaike info criterion		7.664104
Sum squared resid	22847.93	Schwarz criterion		7.732215
Log likelihood	-727.9219	F-statistic		373074.4
Durbin-Watson stat	2.253186	Prob(F-statistic)		0.000000

The fitted values from model  $H_2$  enter significantly in model  $H_1$  and we reject model  $H_1$ .

We must also test model  $H_2$  against model  $H_1$ . Estimate model  $H_1$ , retrieve the fitted values, and estimate model  $H_2$  including the fitted values from model  $H_1$ . The result of this “reverse” test are given by:

Dependent Variable: CS  
Method: Least Squares  
Date: 08/13/97 Time: 16:58  
Sample(adjusted): 1947:2 1994:4  
Included observations: 191 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1427.716	132.0349	-10.81318	0.0000
GDP	5.170543	0.476803	10.84419	0.0000
CS(-1)	0.977296	0.018348	53.26506	0.0000
CS1	-7.292771	0.679043	-10.73978	0.0000
R-squared	0.999833	Mean dependent var		1953.966
Adjusted R-squared	0.999830	S.D. dependent var		848.4387
S.E. of regression	11.05357	Akaike info criterion		7.664104
Sum squared resid	22847.93	Schwarz criterion		7.732215
Log likelihood	-727.9219	F-statistic		373074.4
Durbin-Watson stat	2.253186	Prob(F-statistic)		0.000000

The fitted values are again statistically significant and we reject model  $H_2$ .

In this example, we reject both specifications, against the alternatives, suggesting that another model for the data is needed. It is also possible that we fail to reject both models, in which case the data do not provide enough information to discriminate between the two models.

## Commands

All of the specification and diagnostic tests explained in this chapter are available in command form as views of a named equation. Follow the equation name with a dot and the view name of the test. For example, to carry out the Wald test of whether the third and fourth coefficients of the equation object EQ1 are both equal to zero, type

```
eq1.wald c(3)=0,c(4)=0
```

To carry out the serial correlation LM test of the residuals in equation EQ\_Y up to 4 lags, type

```
eq_y.auto(4)
```

To display the recursive residuals of equation EQM1, type

```
eqm1.rls(r)
```

See the *Command and Programming Reference* for a complete list of commands and options available for equation objects.



## Part IV. Advanced Single Equation Analysis

---

Chapters 16–18 describe EViews tools for the estimation and analysis of advanced single equation models.

- Chapter 16. ARCH and GARCH Estimation—outlines the EViews tools for ARCH and GARCH modeling of the conditional variance, or volatility, of a variable.
- Chapter 17. Discrete and Limited Dependent Variable Models—documents EViews tools for estimating qualitative and limited dependent variable models. EViews provides estimation routines for binary or ordered (probit, logit, gompit), censored or truncated (tobit, etc.), and integer valued (count data).
- Chapter 18. Log Likelihood Estimation—describes techniques for using EViews to estimate the parameters of maximum likelihood models where you may specify the form of the likelihood.

Multiple equation models and forecasting are described in Part V.





## Chapter 16. ARCH and GARCH Estimation

---

Most of the statistical tools in EViews are designed to model the conditional mean of a random variable. The tools described in this chapter differ by modeling the conditional variance, or volatility, of a variable.

There are several reasons that you may want to model and forecast volatility. First, you may need to analyze the risk of holding an asset or the value of an option. Second, forecast confidence intervals may be time-varying, so that more accurate intervals can be obtained by modeling the variance of the errors. Third, more efficient estimators can be obtained if heteroskedasticity in the errors is handled properly.

Autoregressive Conditional Heteroskedasticity (ARCH) models are specifically designed to model and forecast conditional variances. The variance of the dependent variable is modeled as a function of past values of the dependent variable and independent, or exogenous variables.

ARCH models were introduced by Engle (1982) and generalized as GARCH (Generalized ARCH) by Bollerslev (1986). These models are widely used in various branches of econometrics, especially in financial time series analysis. See Bollerslev, Chou, and Kroner (1992) and Bollerslev, Engle, and Nelson (1994) for recent surveys.

In the next section, the basic ARCH model will be described in detail. In subsequent sections, we consider the wide range of specifications available in EViews for modeling volatility. For brevity of discussion, we will use ARCH to refer to both ARCH and GARCH models, except where there is the possibility of confusion.

### The ARCH Specification

In developing an ARCH model, you will have to provide two distinct specifications—one for the conditional mean and one for the conditional variance.

#### The GARCH(1,1) Model

In the standard GARCH(1,1) specification:

$$y_t = x_t' \gamma + \epsilon_t \sigma_t^2 \quad (16.1)$$

$$\sigma_t^2 = \omega + \alpha \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2 \quad (16.2)$$

the mean equation given in (16.1) is written as a function of exogenous variables with an error term. Since  $\sigma_t^2$  is the one-period ahead forecast variance based on past information, it is called the *conditional variance*. The conditional variance equation specified in (2) is a function of three terms:

- The mean:  $\omega$ .
- News about volatility from the previous period, measured as the lag of the squared residual from the mean equation:  $\epsilon_{t-1}^2$  (the ARCH term).
- Last period's forecast variance:  $\sigma_{t-1}^2$  (the GARCH term).

The (1,1) in GARCH(1,1) refers to the presence of a first-order GARCH term (the first term in parentheses) and a first-order ARCH term (the second term in parentheses). An ordinary ARCH model is a special case of a GARCH specification in which there are no lagged forecast variances in the conditional variance equation.

This specification is often interpreted in a financial context, where an agent or trader predicts this period's variance by forming a weighted average of a long term average (the constant), the forecasted variance from last period (the GARCH term), and information about volatility observed in the previous period (the ARCH term). If the asset return was unexpectedly large in either the upward or the downward direction, then the trader will increase the estimate of the variance for the next period. This model is also consistent with the volatility clustering often seen in financial returns data, where large changes in returns are likely to be followed by further large changes.

There are two alternative representations of the variance equation that may aid in the interpretation of the model:

- If we recursively substitute for the lagged variance on the right-hand side of (16.2), we can express the conditional variance as a weighted average of all of the lagged squared residuals:

$$\sigma_t^2 = \frac{\omega}{(1-\beta)} + \alpha \sum_{j=1}^{\infty} \beta^{j-1} \epsilon_{t-j}^2 \quad (16.3)$$

- We see that the GARCH(1,1) variance specification is analogous to the sample variance, but that it down-weights more distant lagged squared errors.
- The error in the squared returns is given by  $v_t = \epsilon_t^2 - \sigma_t^2$ . Substituting for the variances in the variance equation and rearranging terms we can write our model in terms of the errors:

$$\epsilon_t^2 = \omega + (\alpha + \beta)\epsilon_{t-1}^2 + v_t - \beta v_{t-1} \quad (16.4)$$

- Thus, the squared errors follow a heteroskedastic ARMA(1,1) process. The autoregressive root which governs the persistence of volatility shocks is the sum of  $\alpha$  plus  $\beta$ . In many applied settings, this root is very close to unity so that shocks die out rather slowly.

## Regressors in the Variance Equation

Equation (16.2) may be extended to allow for the inclusion of exogenous or predetermined regressors,  $z$ , in the variance equation:

$$\sigma_t^2 = \omega + \alpha \epsilon_t^2 + \beta \sigma_{t-1}^2 + \pi z_t. \quad (16.5)$$

Note that the forecasted variances from this model are not guaranteed to be positive. You may wish to introduce regressors in a form where they are always positive to minimize the possibility that a single, large negative value generates a negative forecasted value. For example, you may want to set

$$z_t = \left| \sqrt{x_t^2} \right|. \quad (16.6)$$

### The ARCH-M Model

The  $x$ 's in equation (16.2) represent exogenous or predetermined variables that are included in the mean equation. If we introduce the conditional variance into the mean equation, we get the ARCH-in-Mean (ARCH-M) model (Engle, Lilien, Robins, 1987):

$$y_t = x_t' \gamma + \sigma_t^2 + \epsilon_t. \quad (16.7)$$

The ARCH-M model is often used in financial applications where the expected return on an asset is related to the expected asset risk. The estimated coefficient on the expected risk is a measure of the risk-return tradeoff.

A variant of the ARCH-M specification uses the conditional standard deviation in place of the conditional variance in Equation (16.7).

### The GARCH(p, q) Model

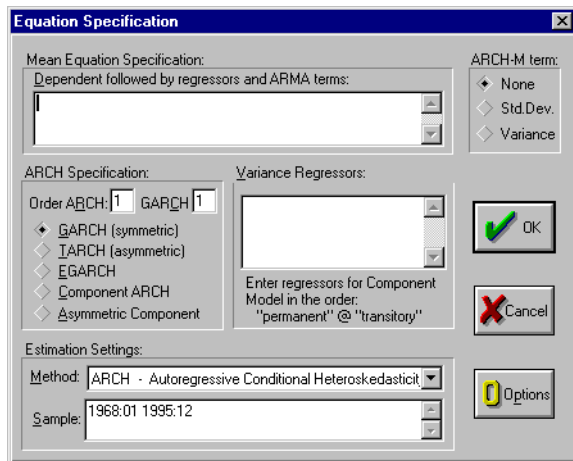
Higher order GARCH models, denoted GARCH( $p, q$ ), can be estimated by choosing either  $p$  or  $q$  greater than 1. The representation of the GARCH( $p, q$ ) variance is

$$\sigma_t^2 = \omega + \sum_{j=1}^p \beta_j \sigma_{t-j}^2 + \sum_{i=1}^q \alpha_i \epsilon_{t-i}^2 \quad (16.8)$$

where  $p$  is the order of the GARCH terms and  $q$  is the order of the ARCH term.

## Estimating ARCH models in EViews

To estimate an ARCH or GARCH model, open the equation specification dialog by selecting **Quick/Estimate Equation...** or by selecting **Object/New Object/Equation....** Select **ARCH** from the method combo box.



You will need to specify both the mean and the variance equations, as well as the estimation technique and sample.

## The Mean Equation

In the dependent variable edit box, you should enter the specification of the mean equation. You can enter the specification in list form by listing the dependent variable followed by the regressors. You should add the C to your specification if you wish to include a constant. If you have a more complex mean specification, you can enter your mean equation using a formula.

If your specification includes an ARCH-M term, you should click on the appropriate radio button in the upper right-hand side of the dialog.

## The Variance Equation

### ARCH Specification

Under the ARCH Specification label, you should choose the number of ARCH and GARCH terms. The default is to estimate with one ARCH and one GARCH term. This is by far the most popular specification.

To estimate the standard GARCH model as described above, click on the GARCH radio button. The other entries describe more complicated variants of the GARCH specification. We discuss each of these models later in the chapter.

### Variance Regressors

In the edit box labeled **Variance Regressors**, you may optionally list variables you wish to include in the variance specification. Note that EViews will always include a constant as a variance regressor so that you do not need to add C to the list.

The distinction between the permanent and transitory regressors is discussed below in “The Component ARCH Model” on page 373 .

## Estimation Options

EViews provides you with access to a number of optional estimation settings. Simply click on the **Options** button and fill out the dialog as required:

### Heteroskedasticity Consistent Covariances

Click on the check box labeled **Heteroskedasticity Consistent Covariance** to compute the quasi-maximum likelihood (QML) covariances and standard errors using the methods described by Bollerslev and Wooldridge (1992).

You should use this option if you suspect that the residuals are not conditionally normally distributed. When the assumption of conditional normality does not hold, the ARCH parameter estimates will still be consistent, provided the mean and variance functions are correctly specified. The estimates of the covariance matrix will not be consistent unless this option is specified, resulting in incorrect standard errors.

Note that the parameter estimates will be unchanged if you select this option; only the estimated covariance matrix will be altered.

### Iterative Estimation Control

The likelihood functions of ARCH models are not always well-behaved so that convergence may not be achieved with the default estimation settings. You can use the options dialog to select the iterative algorithm (Marquardt, BHHH/Gauss-Newton), change starting values, increase the maximum number of iterations, or adjust the convergence criterion.

### Examples

To estimate a standard GARCH(1,1) model with no regressors in the mean and variance equations,

$$\begin{aligned} SP_t &= c + \epsilon_t \sigma_t^2 \\ &= \omega + \alpha \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2 \end{aligned} \quad (16.9)$$

you should enter the various parts of your specification:

- Fill in the Mean Equation Specification edit box as  

$$sp \ c$$
- Enter 1 for the number of ARCH terms, and 1 for the number of GARCH terms, and select **GARCH (symmetric)**.
- Select **None** for the **ARCH-M term**.
- Leave blank the **Variance Regressors** edit box.

To estimate the ARCH(4)-M model

$$\begin{aligned} NYS_t &= \gamma_0 + \gamma_1 DUM_t + \gamma_2 \sigma_t + \epsilon_t \sigma_t^2 \\ &= \omega + \alpha_1 \epsilon_{t-1}^2 + \alpha_2 \epsilon_{t-2}^2 + \alpha_3 \epsilon_{t-3}^2 + \alpha_4 \epsilon_{t-4}^2 + \gamma_3 DUM_t \end{aligned} \quad (16.10)$$

you should fill in the dialog in the following fashion:

- Enter the mean equation specification,  

$$\text{nys dum c}$$
- Enter 4 for the ARCH term and 0 for the GARCH term, and select **GARCH (symmetric)**.
- Select **Std. Dev.** for the **ARCH-M** term.
- Enter DUM in the Variance Regressors edit box.

Once you have filled in the Equation Specification dialog, click **OK** to estimate the model. ARCH models are estimated by the method of maximum likelihood, under the assumption that the errors are conditionally normally distributed. Because the variance appears in a non-linear way in the likelihood function, the likelihood function must be estimated using iterative algorithms. In the status line, you can watch the value of the likelihood as it changes with each iteration. When estimates converge, the parameter estimates and conventional regression statistics are presented in the ARCH object window.

### ARCH Estimation Output

A GARCH(1,1) model fit to the first difference of log daily U.S. Dollar/Japanese Yen exchange rates yielded the following output:

Dependent Variable: R				
Method: ML - ARCH				
Date: 07/30/97 Time: 14:27				
Sample(adjusted): 2 5758				
Included observations: 5757 after adjusting endpoints				
Convergence achieved after 129 iterations				
	Coefficient	Std. Error	z-Statistic	Prob.
C	-0.005172	0.006517	-0.793649	0.4274
Variance Equation				
C	0.016001	0.000304	52.64622	0.0000
ARCH(1)	0.207969	0.003492	59.56064	0.0000
GARCH(1)	0.784126	0.003822	205.1789	0.0000
R-squared	-0.000620	Mean dependent var		-0.020218
Adjusted R-squared	-0.001142	S.D. dependent var		0.604167
S.E. of regression	0.604512	Akaike info criterion		1.596767
Sum squared resid	2102.343	Schwarz criterion		1.601394
Log likelihood	-4592.295	Durbin-Watson stat		1.914825

The output from ARCH estimation is divided into two sections—the upper part provides the standard output for the mean equation, while the lower part, labeled “Variance Equation” contains the coefficients, standard errors,  $z$ -statistics and  $p$ -values for the coefficients of the variance equation. The ARCH parameters correspond to  $\alpha$  in equation (2) and the GARCH parameters to  $\beta$ . The bottom panel of the output presents the standard set of regression statis-

tics using the residuals from the mean equation. Note that measures such as  $R^2$  may not be meaningful if there are no regressors in the mean equation. Here, for example, the  $R^2$  is negative.

In this example, the sum of the ARCH and GARCH coefficients ( $\alpha + \beta$ ) is very close to one, indicating that volatility shocks are quite persistent. This result is often observed in high frequency financial data.

## Working with ARCH Models

Once your model has been estimated, EViews provides a variety of views and procedures for inference and diagnostic checking.

### Views of ARCH Models

**Actual, Fitted, Residual** view displays the residuals in various forms, such as table, graphs, and standardized residuals. You can save the residuals as a named series in your workfile using a procedure (see below).

**Conditional SD Graph** plots the one-step ahead standard deviation  $\sigma_t$  for each observation in the sample. The observation at period  $t$  is the forecast for  $t$  made using information available in  $t - 1$ . You can save the conditional standard deviations as named series in your workfile using a procedure (see below).

**Covariance Matrix** displays the estimated coefficient covariance matrix. Most ARCH models (except ARCH-M models) are block diagonal so that the covariance between the mean coefficients and the variance coefficients is very close to zero. If you include a constant in the mean equation, there will be two C's in the covariance matrix; the first C is the constant of the mean equation, and the second C is the constant of the variance equation.

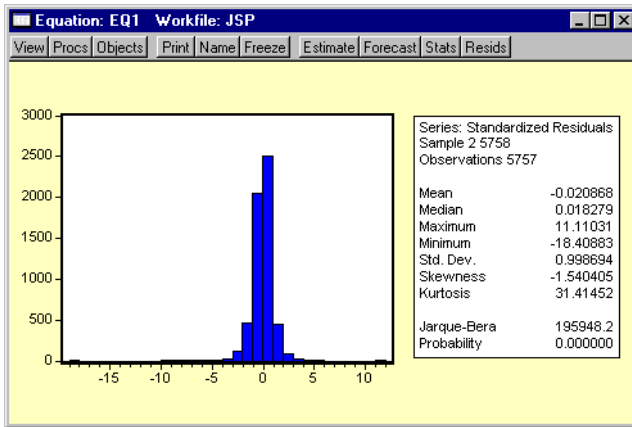
**Coefficient Tests** carries out standard hypothesis tests on the estimated coefficients. See Chapter 14 for details. Note that the likelihood ratio tests are not appropriate under a quasi-maximum likelihood interpretation of your results.

**Residual Tests/Correlogram-Q-statistics** displays the correlogram (autocorrelations and partial autocorrelations) of the standardized residuals. This view can be used to test for remaining serial correlation in the mean equation and to check the specification of the mean equation. If the mean equation is correctly specified, all  $Q$ -statistics should not be significant. See Chapter 7 for an explanation of correlograms and  $Q$ -statistics.

**Residual Tests/Correlogram Squared Residuals** displays the correlogram (autocorrelations and partial autocorrelations) of the squared standardized residuals. This view can be used to test for remaining ARCH in the variance equation and to check the specification of the variance equation. If the variance equation is correctly specified, all  $Q$ -statistics should not be significant. See Chapter 7 for an explanation of correlograms and  $Q$ -statistics. See also **Residual Tests/ARCH LM Test**.

**Residual Tests/Histogram-Normality Test** displays descriptive statistics and a histogram of

the standardized residuals. You can use the Jarque-Bera statistic to test whether the standardized residuals are normally distributed. If the standardized residuals are normally distributed, the Jarque-Bera statistic should not be significant. See Chapter 7 for an explanation of the Jarque-Bera test. For example, the histogram of the standardized residuals from the GARCH(1,1) model fit to the daily exchange rate looks as follows:



The residuals are highly leptokurtic and the Jarque-Bera statistic decisively rejects the hypothesis of normal distribution.

**Residual Tests/ARCH LM Test** carries out Lagrange multiplier tests to test whether the standardized residuals exhibit additional ARCH. If the variance equation is correctly specified, there should be no ARCH left in the standardized residuals. See Chapter 14 for an discussion of ARCH LM tests. See also **Residual Tests/Correlogram Squared Residuals**.

## ARCH Model Procedures

**Make Residual Series** saves the residuals as named series in your workfile. You have the option to save the ordinary residuals,  $\epsilon_t$ , or the standardized residuals,  $\epsilon_t/\sigma_t$ . The residuals will be named RESID1, RESID2, and so on; you can rename the series with the **name** button in the series window.

**Make GARCH Variance Series** saves the conditional variances  $\sigma_t^2$  as named series in your workfile. The conditional variance series will be named GARCH1, GARCH2, and so on. Take the square root to get the conditional standard deviations as displayed by the **View/Conditional SD Graph**.

**Forecast** uses the estimated ARCH model to compute static and dynamic forecasts of the mean, its forecast standard error, and the conditional variance. To save any of these forecasts in your workfile, type a name in the corresponding dialog box. If you choose the **Do graph** option, EViews displays the graphs of the forecasts and two standard deviation bands for the mean forecast. For example, suppose we estimated the following GARCH(1,1)-M model:

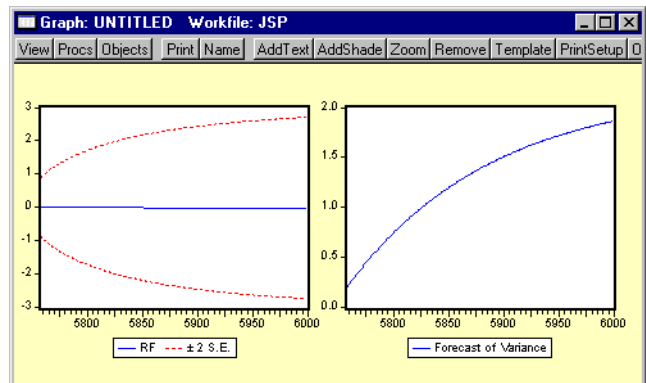


Dependent Variable: R  
 Method: ML - ARCH  
 Date: 07/31/97 Time: 08:30  
 Sample(adjusted): 2 5758  
 Included observations: 5757 after adjusting endpoints  
 Convergence achieved after 205 iterations  
 Bollerslev-Wooldrige robust standard errors & covariance

	Coefficient	Std. Error	z-Statistic	Prob.
SQR(GARCH)	-0.021274	0.038116	-0.558158	0.5767
C	0.004068	0.017355	0.234374	0.8147
Variance Equation				
C	0.016481	0.009140	1.803147	0.0714
ARCH(1)	0.214339	0.051704	4.145479	0.0000
GARCH(1)	0.778094	0.046161	16.85617	0.0000
R-squared	0.000081	Mean dependent var	-0.020218	
Adjusted R-squared	-0.000614	S.D. dependent var	0.604167	
S.E. of regression	0.604352	Akaike info criterion	1.599669	
Sum squared resid	2100.870	Schwarz criterion	1.605452	
Log likelihood	-4599.647	F-statistic	0.116447	
Durbin-Watson stat	1.917673	Prob(F-statistic)	0.976739	

To construct dynamic forecasts of R using this model, click **Forecast** and fill in the Forecast dialog setting the sample after the estimation period. If you choose **Do graph**, the equation view changes to display the forecast results.

The left graph is the forecast of R from the mean equation together with the two standard deviation bands. Although, the point forecasts look constant, they are in fact declining over the forecast period because of the negative sign of the GARCH term in the mean equation (you can check this by looking at the spreadsheet view of the point forecasts.)



The right graph is the forecast of the conditional variance  $\sigma_t^2$ . Since the sum of the ARCH and GARCH terms ( $\alpha + \beta$ ) is close to one, the volatility shocks are persistent, so that the forecasts of the conditional variance converge to the steady state quite slowly.

## Additional Comments

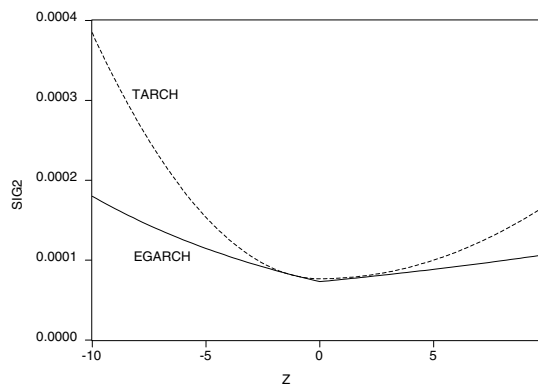
Several of the test results described above are formulated in terms of *standardized residuals*,  $\epsilon_t/\sigma_t$ , which are defined as the conventional mean equation residuals divided by the conditional standard deviation.

If the model is correctly specified, the standardized residuals should be independent, and identically distributed random variables with mean zero and variance one. If the standardized residuals are also normally distributed, then the estimates are maximum likelihood estimates which are asymptotically efficient. However even if the distribution of the residuals is not normal, the estimates are still consistent under quasi-maximum likelihood (QML) assumptions.

To carry out valid inference with QML, you should make certain to use the **Heteroskedasticity Consistent Covariance** option to estimate the standard errors.

## Asymmetric ARCH Models

For equities, it is often observed that downward movements in the market are followed by higher volatilities than upward movements of the same magnitude. To account for this phenomenon, Engle and Ng (1993) describe an asymmetric News Impact Curve with the following shape:



EViews estimates two models that allow for asymmetric shocks to volatility: TARCH and EGARCH.

## The TARCH Model

TARCH or Threshold ARCH was introduced independently by Zakoian (1990) and Glosten, Jagannathan, and Runkle (1993). The specification for the conditional variance is given by

$$\sigma_t^2 = \omega + \alpha \epsilon_{t-1}^2 + \gamma \epsilon_{t-1}^2 d_{t-1} + \beta \sigma_{t-1}^2 \quad (16.11)$$

where  $d_t = 1$  if  $\epsilon_t > 0$ , and 0 otherwise.

In this model, good news ( $\epsilon_t > 0$ ), and bad news ( $\epsilon_t < 0$ ), have differential effects on the conditional variance—good news has an impact of  $\alpha$ , while bad news has an impact of  $(\alpha + \gamma)$ . If  $\gamma > 0$  we say that the *leverage effect* exists. If  $\gamma \neq 0$ , the news impact is asymmetric.

For higher order specifications of the TARCH model, EViews estimates

$$\sigma_t^2 = \omega + \sum_{i=1}^q \alpha_i \epsilon_{t-i}^2 + \gamma \epsilon_{t-1}^2 d_{t-1} + \sum_{j=1}^p \beta_j \sigma_{t-j}^2 \quad (16.12)$$

To estimate this model, specify your ARCH model in the usual fashion, but instead of selecting a GARCH specification, you should click on the **TARCH (asymmetric)** radio button listed under **ARCH Specification**.

The TARCH(1,1) model fitted to the daily exchange rate returns gives:

Dependent Variable: R				
Method: ML - ARCH				
Date: 07/31/97 Time: 11:04				
Sample(adjusted): 2 5758				
Included observations: 5757 after adjusting endpoints				
Convergence achieved after 169 iterations				
Bollerslev-Wooldrige robust standard errors & covariance				
	Coefficient	Std. Error	z-Statistic	Prob.
C	-0.005342	0.005782	-0.923938	0.3555
Variance Equation				
C	0.015829	0.008728	1.813624	0.0697
ARCH(1)	0.206057	0.056741	3.631574	0.0003
(RESID<0)*ARCH(1)	0.004405	0.050454	0.087312	0.9304
GARCH(1)	0.784364	0.045053	17.40984	0.0000
R-squared	-0.000606	Mean dependent var	-0.020218	
Adjusted R-squared	-0.001302	S.D. dependent var	0.604167	
S.E. of regression	0.604560	Akaike info criterion	1.597089	
Sum squared resid	2102.314	Schwarz criterion	1.602872	
Log likelihood	-4592.221	Durbin-Watson stat	1.914852	

The leverage effect term,  $\gamma$ , represented by (RESID < 0)\*ARCH(1) in the output, is not significantly positive (even with a one-sided test) so there appears to be no asymmetric effect. Note that we have used quasi-likelihood robust standard errors since the residuals are highly leptokurtic.

Note that when forecasting with this model, EViews assumes that the distribution of the residuals is symmetric so that  $d = 1$  half of the time. Since we cannot identify when this occurs, we arbitrarily set  $d = 0.5$  for all observations.

### The EGARCH Model

The EGARCH or Exponential GARCH model was proposed by Nelson (1991). The specification for the conditional variance is

$$\log \sigma_t^2 = \omega + \beta \log \sigma_{t-1}^2 + \alpha \left| \frac{\epsilon_{t-1}}{\sigma_{t-1}} \right| + \gamma \frac{\epsilon_{t-1}}{\sigma_{t-1}}. \quad (16.13)$$

Note that the left-hand side is the *log* of the conditional variance. This implies that the leverage effect is exponential, rather than quadratic, and that forecasts of the conditional variance are guaranteed to be nonnegative. The presence of leverage effects can be tested by the hypothesis that  $\gamma > 0$ . The impact is asymmetric if  $\gamma \neq 0$ .

There are two differences between the EViews specification of the EGARCH model and the original Nelson model. First, Nelson assumes that the  $\epsilon$  follows a generalized error distribution, while EViews assumes normally distributed errors. Second, Nelson's specification for the log conditional variances differs slightly from the specification above:

$$\log \sigma_t^2 = \omega + \beta \log \sigma_{t-1}^2 + \alpha \left| \frac{\epsilon_{t-1}}{\sigma_{t-1}} - \sqrt{\frac{2}{\pi}} \right| + \gamma \frac{\epsilon_{t-1}}{\sigma_{t-1}} \quad (16.14)$$

Estimating this model under the assumption of normal errors will yield identical estimates to those reported by EViews except for the intercept term  $\omega$ , which will differ by  $\alpha \sqrt{2/\pi}$ .

$$\log \sigma_t^2 = \omega + \sum_{i=1}^p \beta_i \log \sigma_{t-i}^2 + \sum_{j=1}^q \left( \alpha_j \left| \frac{\epsilon_{t-j}}{\sigma_{t-j}} - \sqrt{\frac{2}{\pi}} \right| + \gamma_j \frac{\epsilon_{t-j}}{\sigma_{t-j}} \right) \quad (16.15)$$

To estimate an EGARCH model, simply select the EGARCH radio button under the ARCH specification settings. Applying this method to the daily bond futures data gives the following results:

Dependent Variable: BF  
 Method: ML - ARCH  
 Date: 07/31/97 Time: 14:30  
 Sample: 1 2114  
 Included observations: 2114  
 Convergence achieved after 44 iterations  
 Bollerslev-Wooldrige robust standard errors & covariance

	Coefficient	Std. Error	z-Statistic	Prob.
C	8.42E-05	0.000134	0.625786	0.5315
Variance Equation				
C	-0.136413	0.036840	-3.702907	0.0002
IRESI/SQR[GARCH](1)	0.064330	0.015611	4.120901	0.0000
RES/SQR[GARCH](1)	-0.025932	0.011330	-2.288767	0.0221
EGARCH(1)	0.991272	0.003095	320.2354	0.0000
R-squared	0.000000	Mean dependent var	8.78E-05	
Adjusted R-squared	-0.001897	S.D. dependent var	0.007017	
S.E. of regression	0.007024	Akaike info criterion	-7.221349	
Sum squared resid	0.104050	Schwarz criterion	-7.207971	
Log likelihood	7637.966	Durbin-Watson stat	1.934669	

The leverage effect term,  $\gamma$ , denoted as RES/SQR[GARCH](1) in the output, is negative and statistically different from zero, indicating the existence of the leverage effect in future bonds returns during the sample period.

## Plotting the Estimated News Impact Curve

What does the estimated news impact curve look like? Here we illustrate one method of using EViews to plot this curve. Our goal is to plot the volatility  $\sigma^2$ , against the impact  $z = \epsilon/\sigma$ , where

$$\log \sigma_t^2 = \hat{\omega} + \hat{\beta} \log \sigma_{t-1}^2 + \hat{\alpha} |z_{t-1}| + \hat{\gamma} z_{t-1} \quad (16.16)$$

We will fix last period's volatility  $\sigma_{t-1}^2$  to the median of the estimated conditional variance series and estimate the one-period impact, conditional on last period's volatility.

Suppose you estimated the above EGARCH model in an undated workfile and saved the equation as an object named EQ1. First, generate the conditional variance series by clicking **Procs/Make GARCH Variance Series**. Next, store the median by entering the following command in the command window:

```
scalar med = @median(garch1)
```

where GARCH1 is the name of the conditional variance series.

Generate the  $z$  series, which will be the  $x$ -axis of the news impact curve, using the following commands:

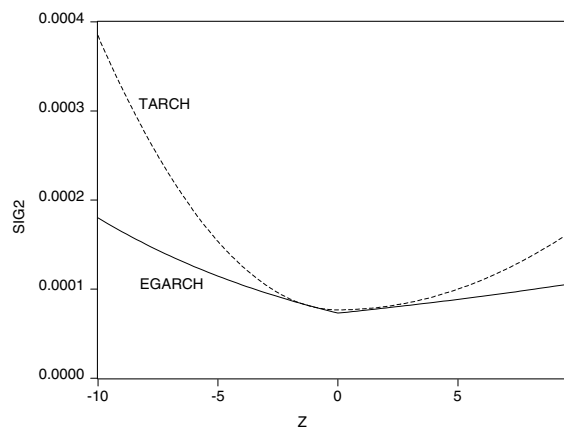
```
smp1 1 100
series z = -10 + @trend(1)*20/100
```

This generates an equispaced  $z$  series between  $-10$  and  $10$ . Then generate the  $\sigma^2$  series by the command

```
series log(sig2) = eq1.c(2) + eq1.c(5)*log(med) +
eq1.c(3)*abs(z) + eq1.c(4)*z
```

where SIG2 is the name for the  $\sigma^2$  series. Note that EViews will automatically create the series SIG2 from the log specification.

Finally, highlight the two series Z and SIG2 (in order to get the axis right), double click, **Open Group**, and **View/Graph/XY line**. Below we show a customized graph depicting the estimated news impact curves from TARCH and EGARCH models fitted to the daily futures return:



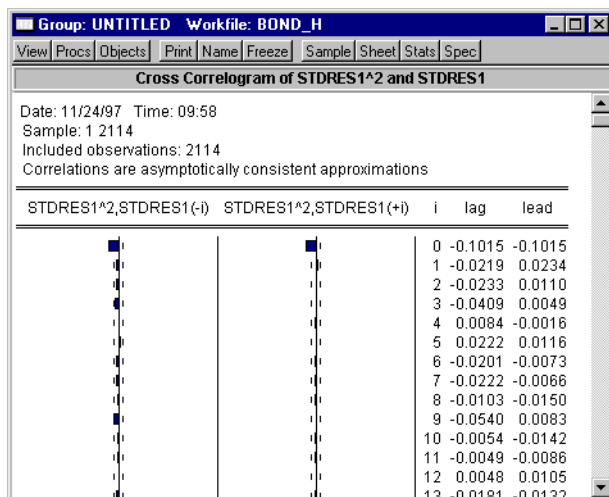
### Checking for Asymmetry Using Cross Correlations

Estimating the TARCH and EGARCH models and testing the significance of the asymmetric terms is one way to test for asymmetric effects. Alternatively, we can look at the cross correlation between the squared standardized residuals  $z_t^2$  and lagged standardized residuals  $z_{t-k}$ . These cross correlations should be zero for a symmetric GARCH model and negative for a TARCH or EGARCH asymmetric model.

To display the cross correlogram between  $z_t^2$  and  $z_{t-k}$ , first save the standardized residuals from the estimated model in a series named STDRES1. Then type the command:

```
show stdres1^2 stdres1
```

where RESID1 is the name of the standardized residuals. Then **View/Cross Correlation (2)** ... and specify the lags to display the cross correlogram. The cross correlogram for the simple GARCH(1,1) model fit to future bonds returns is displayed below.



Because of the large sample size, the approximate two standard error bands (the dotted lines) are very tight around the estimated values. Note that the cross correlation only picks up linear associations between the two series and may miss nonlinear dependence between the two series.

## The Component ARCH Model

The conditional variance in the GARCH(1,1) model:

$$\sigma_t^2 = \bar{\omega} + \alpha(\epsilon_{t-1}^2 - \bar{\omega}) + \beta(\sigma_{t-1}^2 - \bar{\omega}) \quad (16.17)$$

shows mean reversion to  $\bar{\omega}$  which is a constant for all time. By contrast, the component model allows mean reversion to a varying level  $q_t$ , modeled as:

$$\begin{aligned} \sigma_t^2 - q_t &= \bar{\omega} + \alpha(\epsilon_{t-1}^2 - \bar{\omega}) + \beta(\sigma_{t-1}^2 - \bar{\omega}) \\ q_t &= \omega + \rho(q_{t-1} - \omega) + \phi(\epsilon_{t-1}^2 - \sigma_{t-1}^2) \end{aligned} \quad (16.18)$$

Here  $\sigma_t^2$  is still the volatility, while  $q_t$  takes the place of  $\omega$  and is the time varying long run volatility. The first equation describes the transitory component,  $\sigma_t^2 - q_t$ , which converges to zero with powers of  $(\alpha + \beta)$ . The second equation describes the long run component  $q_t$ , which converges to  $\omega$  with powers of  $\rho$ . Typically  $\rho$  is between .99 and 1 so that  $q_t$  approaches  $\omega$  very slowly. We can combine the transitory and permanent equations and write

$$\begin{aligned}\sigma_t^2 = & (1 - \alpha - \beta)(1 - \rho)\omega + (\alpha + \phi)\epsilon_{t-1}^2 - (\alpha\rho + (\alpha + \beta)\phi)\epsilon_{t-2}^2 \\ & + (\beta - \phi)\sigma_{t-1}^2 - (\beta\rho - (\alpha + \beta)\phi)\sigma_{t-2}^2\end{aligned}\quad (16.19)$$

which shows that the component model is a (nonlinear) restricted GARCH(2,2) model.

You can include exogenous variables in the conditional variance equation of component models, either in the permanent or transitory equation (or both). The variables in the transitory equation will have an impact on the short run movements in volatility, while the variables in the permanent equation will affect the long run levels of volatility.

The **Asymmetric Component** option in the Equation Specification dialog combines the component model with the asymmetric TARCh model. This specification introduces asymmetric effects in the transitory equation and estimates models of the form:

$$\begin{aligned}y_t &= x_t' \pi + \epsilon_t \\ q_t &= \omega + \rho(q_{t-1} - \omega) + \phi(\epsilon_{t-1}^2 - \sigma_{t-1}^2) + \theta_1 z_{1t} \\ \sigma_t^2 - q_t &= \alpha\epsilon_{t-1}^2 - q_{t-1} + \gamma\epsilon_{t-1}^2 - q_{t-1}d_{t-1} + \beta\sigma_{t-1}^2 - q_{t-1} + \theta_2 z_{2t}\end{aligned}\quad (16.20)$$

where  $z$ 's are the exogenous variables and  $d$  is the dummy variable indicating negative shocks.  $\gamma > 0$  indicates the presence of transitory leverage effects in the conditional variance.

## Estimating and Interpreting Models in EViews

To estimate component models in EViews, choose either the **Component ARCH** or **Asymmetric Component** option in the Equation Specification dialog. To include exogenous regressors in the variance equation, type the names of the regressors in the **Variance Regressors** box in the following order: first list the series names to include in the permanent equation, followed by an @ sign, and then list the series names to include in the transitory equation. For example, to include HOL in the permanent equation and JAN, END in the transitory equation, type

```
hol @ jan end
```

and to include JAN in the transitory equation, type

```
@ jan
```

The estimation results of fitting the component model to the same bond futures return data is shown below:



Dependent Variable: BF  
 Method: ML - ARCH  
 Date: 08/01/97 Time: 07:25  
 Sample: 1 2114  
 Included observations: 2114  
 Convergence achieved after 27 iterations  
 Bollerslev-Wooldrige robust standard errors & covariance

	Coefficient	Std. Error	z-Statistic	Prob.
C	0.000127	0.000134	0.946855	0.3437
Variance Equation				
Perm: C	4.39E-05	8.12E-06	5.411813	0.0000
Perm: [Q-C]	0.990484	0.004136	239.4579	0.0000
Perm: [ARCH-GARCH]	0.031646	0.008339	3.794929	0.0001
Tran: [ARCH-Q]	-0.007779	0.017883	-0.435004	0.6636
Tran: [GARCH-Q]	-0.844212	0.445347	-1.895627	0.0580
R-squared	-0.000030	Mean dependent var	8.78E-05	
Adjusted R-squared	-0.002402	S.D. dependent var	0.007017	
S.E. of regression	0.007026	Akaike info criterion	-7.212442	
Sum squared resid	0.104053	Schwarz criterion	-7.196388	
Log likelihood	7629.551	Durbin-Watson stat	1.934611	

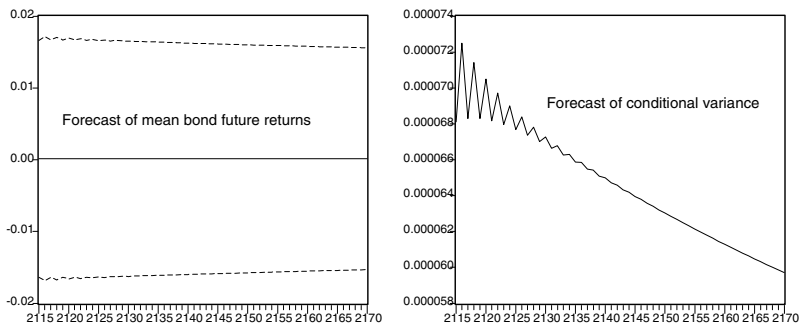
The coefficients labeled "Perm:" are the coefficients for the permanent equation and Trans: for the transitory equation. The estimate of the persistence in the long run component is  $\hat{\rho} = 0.99$ , indicating that the long run component converges very slowly to the steady state. Short run volatility is governed by powers of  $(\hat{\alpha} + \hat{\beta}) = -0.85$ , which is marginally significant. To test the hypothesis  $(\alpha + \beta) = 0$ , choose **View/Coefficient Tests/Wald...** and type

$$c(5) + c(6) = 0$$

in the Wald Test dialog box. The result is given by:

Wald Test:			
Equation: EQ_C1			
Null Hypothesis: C(5)+C(6)=0			
F-statistic	3.856177	Probability	0.049694
Chi-square	3.856177	Probability	0.049563

To forecast from the component model, click **Forecast**, enter 2115 2170 as the forecast interval, and make certain that **Do Graph** is selected. Click **OK**. Here we have copy-and-pasted the customized graph output into our word processor

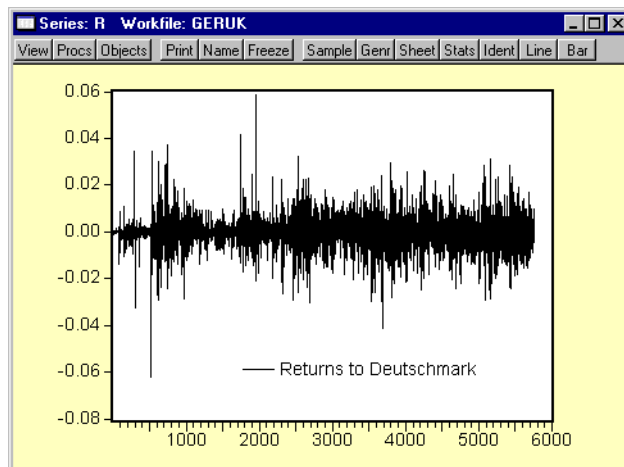


Notice that the volatility forecast from the component model need not be monotonic. In this example, there is a short run fluctuating component which dies out in the long run.

## Examples

### Modeling the volatility in the US Dollar/Deutschmark exchange rates

As an illustration of ARCH modeling in EViews, we use daily US Dollar/Deutschmark exchange rate from 1971 to 1993. The dependent variable is the daily nominal return,  $r_t = \log(s_t/s_{t-1})$ , where  $s$  is the spot rate. The return series clearly shows volatility clustering, especially early in the sample.



We start from a general GARCH(1,1)-M model of the following form:

$$r_t = \pi_0 + \pi_1 \sigma_t^2 + \epsilon_t + \theta \epsilon_{t-1} \quad (16.21)$$

$$\sigma_t^2 = \omega + \alpha \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2 + \gamma MON_t$$

The mean equation includes the conditional variance and the errors are MA(1). The MON series in the conditional variance equation is a dummy variable for Monday, which is meant to capture the non-trading period effect during the weekend. The results are given below:

Dependent Variable: R

Method: ML - ARCH

Date: 08/01/97 Time: 11:23

Sample(adjusted): 2 5764

Included observations: 5763 after adjusting endpoints

Convergence achieved after 130 iterations

Bollerslev-Wooldrige robust standard errors & covariance

Backcast: 1

	Coefficient	Std. Error	z-Statistic	Prob.
GARCH	-0.879360	2.762081	-0.318369	0.7502
C	-7.22E-05	0.000101	-0.712321	0.4763
MA(1)	0.020390	0.015824	1.288485	0.1976
Variance Equation				
C	1.85E-06	1.81E-06	1.019493	0.3080
ARCH(1)	0.120933	0.013406	9.020821	0.0000
GARCH(1)	0.866180	0.012901	67.13988	0.0000
MON	-4.35E-06	7.36E-06	-0.590873	0.5546
R-squared	0.001059	Mean dependent var	-0.000128	
Adjusted R-squared	0.000018	S.D. dependent var	0.006628	
S.E. of regression	0.006628	Akaike info criterion	-7.404128	
Sum squared resid	0.252876	Schwarz criterion	-7.396039	
Log likelihood	21341.99	F-statistic	1.017004	
Durbin-Watson stat	1.973174	Prob(F-statistic)	0.411982	
Inverted MA Roots	-.02			

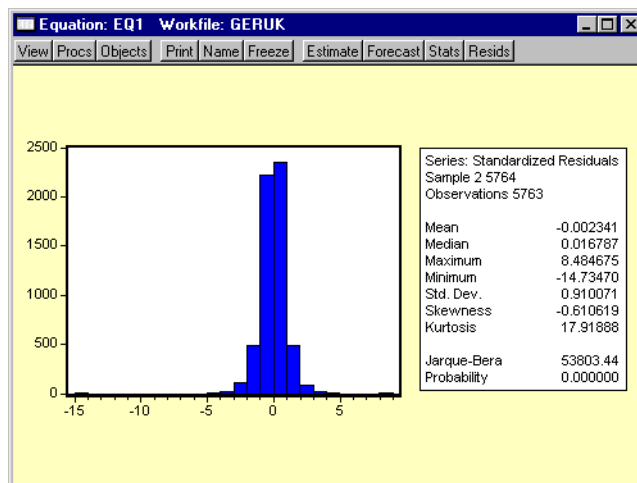
To check whether there are any ARCH effects left in the residuals, click **View/Residual Tests/ARCH LM Test...** and specify the order to test. The top part of the output from testing up to an ARCH(7) is given by

ARCH Test:

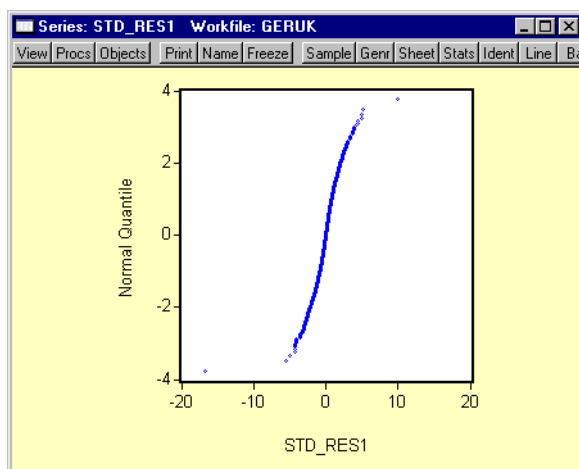
F-statistic	0.174264	Probability	0.990453
Obs*R-squared	1.221284	Probability	0.990429

indicating that there does not appear to be any ARCH up to order 7.

To check whether the distribution of the standardized residuals look normal, click **View/Residual Tests/Histogram**:



The residuals are highly leptokurtic and the Jarque-Bera test decisively rejects the normal distribution. An alternative way to check the distribution of the residuals is to plot the quantiles. First save the residuals by clicking **Procs/Make Residual Series...** and choose **standardized**. Then in the residual series window choose **View/Distribution Graphs/Quantile-Quantile...** and plot against **quantiles of the normal distribution**:



If the residuals are normally distributed, the QQ-plots should lie on a straight line; see Chapter 9 for details on QQ-plots. The plot shows that it is primarily a few large outliers that are driving the departure from normality. You can also experiment with other distributions. For example, in the ARCH literature, the  $t$ -distribution is commonly used to capture the thick tails in the distribution. Although there is no option for the  $t$ -distribution in the **Quantile-Quantile** view, you can simu-

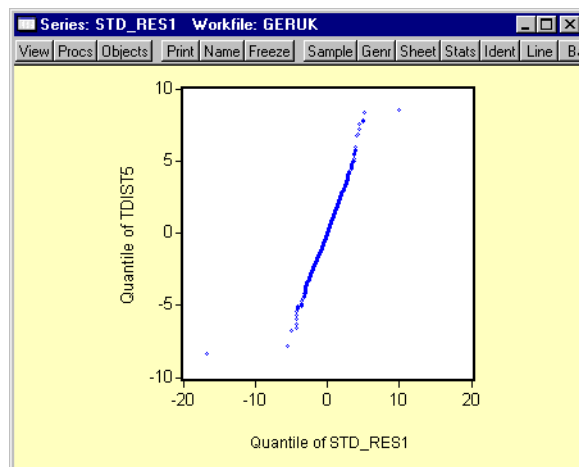
late a draw from a  $t$ -distribution and check whether the distribution matches with the residuals. The command

```
series tdist2 = @qtdist(rnd,5)
```

simulates a random draw from the  $t$ -distribution with 5 degrees of freedom. Then in the QQ Plot dialog box choose **Series or Group** and type the name of the series (in this case `tdist2`).

The QQ-plot now lies mostly on a straight line, except for two outlier observations.

The departure from normality of the residuals suggests using the robust standard errors option: click **Options** in the Equation Specification dialog and click **Heteroskedasticity Consistent Covariance** option.



## Commands

The ARCH procedure estimates GARCH models:

```
equation eq1.arch(1,1) sp500 c
```

estimates a GARCH(1,1) model with only a constant in the conditional mean equation.

```
eq1.arch(1,1) sp500 c @ d_mon
```

estimates a GARCH(1,1) model with a constant in the conditional mean equation and `D_MON` in the conditional variance equation.

```
eq1.makegarch garch1
```

stores the estimated conditional variance series from `EQ1` as a series named `GARCH1`.

See the *Command and Programming Reference* for a complete list of commands and options available in ARCH estimation.

## Technical Details

ARCH models in EViews are estimated by the method of maximum likelihood under the assumption that the errors are conditionally normally distributed.

For example, for the GARCH(1,1) model, the contribution to the log likelihood from observation  $t$  is

$$l_t = \log \sqrt{2\pi} + \log \sqrt{\sigma_t^2} - (y_t - x_t' \pi)^2 / 2\sigma_t^2, \quad (16.22)$$

where  $\sigma_t^2 = \omega + \alpha(y_{t-1} - x_{t-1}' \pi)^2 + \beta\sigma_{t-1}^2$ . Since the variance specification enters the likelihood function nonlinearly, the likelihood function must be estimated using iterative algorithms. EViews offers you the choice between Marquardt (the default algorithm) and BHHH/Gauss-Newton. While the former should be sufficient for most needs, you may want to try BHHH if you encounter convergence problems; simply select the BHHH option in the Options menu of the estimation dialog.

As with other iterative procedures, starting values are required. EViews will supply its own starting values to ARCH procedures unless you use the option to specify your own. To obtain starting values, EViews runs a preliminary OLS regression and sets the initial values as

$$\sigma_0^2 = \epsilon_0^2 = \lambda^T s^2 + (1 - \lambda) \sum_{j=0}^{T-1} \lambda^{T-j-1} \epsilon_{T-j}^2, \quad (16.23)$$

where  $\lambda = 0.7$ ,  $\hat{\epsilon}$  is the OLS residuals from the mean equation, and  $s^2$  its variance. Using the Options dialog, you can set starting values to various fractions of the OLS starting values, or you can specify them yourself in the usual fashion.

# Chapter 17. Discrete and Limited Dependent Variable Models

---

The regression methods described in Chapters 11 and 12 require that the dependent variable be observed on a continuous and unrestricted scale. It is quite common, however, for this condition to be violated, resulting in a non-continuous, or a limited dependent variable. We will distinguish between three types of these variables:

- qualitative (observed on a discrete or ordinal scale)
- censored or truncated
- integer valued

In this chapter we discuss estimation methods for several qualitative and limited dependent variable models. EViews provides estimation routines for binary or ordered (probit, logit, gompit), censored or truncated (tobit, etc.), and integer valued (count data) models.

Standard introductory discussion for the models presented in this chapter may be found in Greene (1997), Johnston and DiNardo (1997), and Maddala (1983). Wooldridge (1996) provides an excellent reference for quasi-likelihood methods and count models.

## Binary Dependent Variable Models

In this class of models, the dependent variable,  $y$ , may take on only two values— $y$  might be a dummy variable representing the occurrence of an event, or a choice between two alternatives. For example, you may be interested in modeling the employment status of each individual in your sample (whether employed or not). The individuals differ in age, educational attainment, race, marital status, and other observable characteristics, which we denote as  $x$ . The goal is to quantify the relationship between the individual characteristics and the probability of being employed.

### Theory

Suppose that a binary dependent variable,  $y$ , takes on values of zero and one. A simple linear regression of  $y$  on  $x$  is not appropriate, since among other things, the implied model of the conditional mean places inappropriate restrictions on the residuals of the model. Furthermore, the fitted value of  $y$  from a simple linear regression is not restricted to lie between zero and one.

Instead, we adopt a specification that is designed to handle the specific requirements of binary dependent variables. Suppose that we model the probability of observing a value of one as:

$$\Pr(y_i = 1|x_i, \beta) = 1 - F(-x_i'\beta), \quad (17.1)$$

where  $F$  is a continuous, strictly increasing function that takes a real value and returns a value ranging from zero to one. The choice of the function  $F$  determines the type of binary model. It follows that

$$\Pr(y_i = 0|x_i, \beta) = F(-x_i'\beta). \quad (17.2)$$

Given such a specification, we can estimate the parameters of this model using the method of maximum likelihood. The likelihood function is given by:

$$l(\beta) = \sum_{i=0}^n y_i \log(1 - F(-x_i'\beta)) + (1 - y_i) \log(F(-x_i'\beta)). \quad (17.3)$$

The first order conditions for this likelihood are nonlinear so that obtaining parameter estimates requires an iterative solution. By default, EViews uses a second derivative method for iteration and computation of the covariance matrix of the parameter estimates. As discussed below, EViews allows you to override these defaults using the Options dialog (see Appendix E, page 591, for additional details on the estimation methods).

There are two alternative interpretations of this specification that are of interest. First, the binary model is often motivated as a latent variables specification. Suppose that there is an unobserved latent variable  $y_i^*$  that is linearly related to  $x$ :

$$y_i^* = x_i'\beta + u_i \quad (17.4)$$

where  $u_i$  is a random disturbance. Then the observed dependent variable is determined by whether  $y_i^*$  exceeds a threshold value:

$$y_i = \begin{cases} 1 & \text{if } y_i^* > 0 \\ 0 & \text{if } y_i^* \leq 0 \end{cases} \quad (17.5)$$

In this case, the threshold is set to zero, but the choice of a threshold value is irrelevant, so long as a constant term is included in  $x_i$ . Then

$$\Pr(y_i = 1|x_i, \beta) = \Pr(y_i^* > 0) = \Pr(x_i'\beta + u_i > 0) = F_u(-x_i'\beta) \quad (17.6)$$

where  $F_u$  is the cumulative distribution function of  $u$ . Common models include probit (standard normal), logit (logistic), and gompit (extreme value) specifications for the  $F$  function.

In principle, the coding of the two numerical values of  $y$  is not critical since each of the binary responses only represents an event. Nevertheless, EViews requires that you code  $y$  as a zero-one



variable. This restriction yields a number of advantages. For one, coding the variable in this fashion implies that expected value of  $y$  is simply the probability that  $y = 1$ :

$$\begin{aligned} E(y_i|x_i, \beta) &= 1 \cdot \Pr(y_i = 1|x_i, \beta) + 0 \cdot \Pr(y_i = 0|x_i, \beta) \\ &= \Pr(y_i = 1|x_i, \beta) \end{aligned} \quad (17.7)$$

Thus, we have the second interpretation of the binary specification: as a conditional mean specification. It follows that we can write the binary model as a regression model:

$$y_i = (1 - F(-x_i' \beta)) + \epsilon_i, \quad (17.8)$$

where  $\epsilon_i$  is a residual representing the deviation of the binary  $y_i$  from its conditional mean. Then

$$\begin{aligned} E(\epsilon_i|x_i, \beta) &= 0 \\ \text{var}(\epsilon_i|x_i, \beta) &= F(-x_i' \beta)(1 - F(-x_i' \beta)) \end{aligned} \quad (17.9)$$

We will use the conditional mean interpretation in our discussion of binary model residuals below.

## Estimating Binary Models in EViews

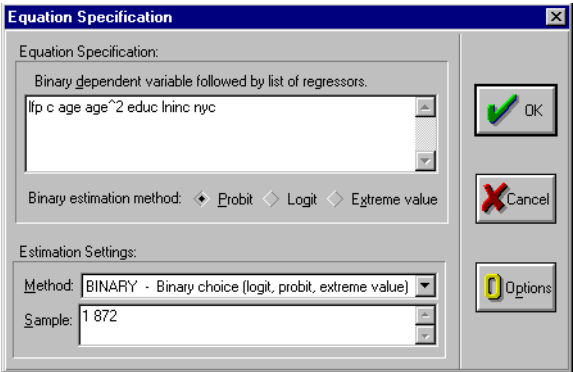
To estimate a binary dependent variable model, choose **Objects/New Object...** from the main menu and select the **Equation** object from the main menu. From the Equation Specification dialog, select the **BINARY** estimation method. The dialog will change to reflect your choice.

There are two parts to the binary model specification. First, in the **Equation Specification** field, you should type the name of the binary dependent variable followed by a list of regressors. You may not enter an explicit equation since binary estimation only supports specification by list. Next, select from among the three distributions for your error term:

Probit	$\Pr(y_i = 1 x_i, \beta) = 1 - \Phi(-x_i' \beta) = \Phi(x_i' \beta)$ <p>where <math>\Phi</math> is the cumulative distribution function of the standard normal distribution.</p>
--------	--

Logit	$\Pr(y_i = 1 x_i, \beta) = 1 - (e^{-x_i'\beta} / (1 + e^{-x_i'\beta}))$ $= e^{x_i'\beta} / (1 + e^{x_i'\beta})$ <p>which is based upon the cumulative distribution function for the logistic distribution.</p>
Extreme value (Gompit)	$\Pr(y_i = 1 x_i, \beta) = 1 - (1 - \exp(-e^{-x_i'\beta}))$ $= \exp(-e^{-x_i'\beta})$ <p>which is based upon the CDF for the Type-I extreme value distribution. Note that this distribution is skewed.</p>

For example, consider the probit specification given by:



where the series labels are given by

<input checked="" type="checkbox"/>	age	09/10/97 17:26	
<input checked="" type="checkbox"/>	c	09/10/97 17:26	
<input checked="" type="checkbox"/>	educ	09/10/97 17:27	Display Name: years of education
<input checked="" type="checkbox"/>	lfp	09/10/97 17:27	Display Name: labor force participation dummy
<input checked="" type="checkbox"/>	lninc	09/10/97 17:27	Display Name: log of nonlabor income
<input checked="" type="checkbox"/>	nyc	09/10/97 17:27	Display Name: number of children below 7 years old
<input checked="" type="checkbox"/>	resid	09/10/97 17:26	

These example data are taken from Gerfin, M., “Parametric and semiparametric estimation of the binary response model of labour market participation,” *Journal of Applied Econometrics*, Vol. 11, No. 3, 1996, 321-340.

Once you have specified the model, click OK. EViews estimates the parameters of the model using iterative procedures, and will display information in the status line. EViews requires that the dependent variable be coded with the values zero-one with all other observations dropped from the estimation.

Following estimation, EViews displays results in the equation window. The top part of the estimation output is given by:

Dependent Variable: LFP				
Method: ML - Binary Probit				
Date: 10/01/97 Time: 23:37				
Sample: 1 872				
Included observations: 872				
Convergence achieved after 5 iterations				
Covariance matrix computed using second derivatives				
Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	5.530194	1.372378	4.029645	0.0001
AGE	1.822439	0.359982	5.062580	0.0000
AGE^2	-0.261397	0.043416	-6.020754	0.0000
EDUC	-0.019168	0.016078	-1.192156	0.2332
LNINC	-0.760094	0.131391	-5.784960	0.0000
NYC	-0.560133	0.091661	-6.110897	0.0000

The header contains basic information regarding the estimation technique (ML for maximum likelihood) and the sample used in estimation, as well as information on the number of iterations required for convergence, and on the method used to compute the coefficient covariance matrix.

Displayed next are the coefficient estimates, asymptotic standard errors,  $z$ -statistics and corresponding  $p$ -values.

Interpretation of the coefficient values is complicated by the fact that estimated coefficients from a binary model cannot be interpreted as the marginal effect on the dependent variable. The marginal effect of  $x_j$  on the conditional probability is given by:

$$\frac{\partial E(y_i|x_i, \beta)}{\partial x_{ij}} = -f(-x_i'\beta)\beta_j, \quad (17.10)$$

where  $f(x) = dF(x)/dx$  is the density function corresponding to  $F$ . Note that  $\beta_j$  is weighted by a factor  $f$  that depends on the values of *all* of the regressors in  $x$ . The *direction* of the effect of a change in  $x_j$  depends only on the sign of the  $\beta_j$  coefficient. Positive values of  $\beta_j$  imply that increasing  $x_j$  will increase the probability of the response; negative values imply the opposite.

Under “Forecast” on page 393 below, we show you how to use EViews to compute these marginal effects.

An alternative interpretation of the coefficients results from noting that the ratios of coefficients provide a measure of the relative changes in the probabilities:

$$\frac{\beta_j}{\beta_k} = \frac{\partial E(y_i|x_i, \beta)/\partial x_{ij}}{\partial E(y_i|x_i, \beta)/\partial x_{ik}}. \quad (17.11)$$

In addition to the summary statistics of the dependent variable, EViews also presents the following summary statistics:

Mean dependent var	0.459862	S.D. dependent var	0.498672
S.E. of regression	0.460668	Akaike info criterion	1.232549
Sum squared resid	183.7781	Schwarz criterion	1.265376
Log likelihood	-531.3915	Hannan-Quinn criter.	1.245108
Restr. log likelihood	-601.6117	Avg. log likelihood	-0.609394
LR statistic (5 df)	140.4404	McFadden R-squared	0.116720
Probability(LR stat)	0.000000		
Obs with Dep=0	471	Total obs	872
Obs with Dep=1	401		

First, there are several familiar summary descriptive statistics: the mean and standard deviation of the dependent variable, standard error of the regression, and the sum of the squared residuals. The latter two measures are computed in the usual fashion using the residuals:

$$e_i = y_i - E(y_i | x_i, \hat{\beta}) = y_i - (1 - F(-x_i' \hat{\beta})) \quad (17.12)$$

Additionally, there are several likelihood based statistics:

**Log likelihood** is the maximized value of the log likelihood function  $l(\hat{\beta})$ .

**Avg. log likelihood** is the log likelihood  $l(\hat{\beta})$  divided by the number of observations  $n$ .

**Restr. log likelihood** is the maximized log likelihood value, when all slope coefficients (except for the constant term) are restricted to zero,  $l(\tilde{\beta})$

The **LR statistic** tests the joint null hypothesis that all slope coefficients except the constant are zero and is computed as  $-2(l(\tilde{\beta}) - l(\hat{\beta}))$ . This statistic, which is only reported when you included a constant in your specification, is used to test the overall significance of the model. The number in parentheses is the degrees of freedom, which is the number of restrictions under test.

**Probability(LR stat)** is the  $p$ -value of the LR test statistic. Under the null hypothesis, the LR test statistic is asymptotically distributed as a  $\chi^2$  variable, with degrees of freedom equal to the number of restrictions under test.

**McFadden R-squared** is the likelihood ratio index computed as  $1 - l(\tilde{\beta})/l(\hat{\beta})$ , where  $l(\tilde{\beta})$  is the restricted log likelihood. As the name suggests, this is an analog to the  $R^2$  reported in linear regression models. It has the property that it always lies between zero and one.

The information criteria are computed as:

<b>Akaike info criterion (AIC)</b>	$-2l(\hat{\beta})/n + 2k/n$
<b>Schwarz criterion (SC)</b>	$-2l(\hat{\beta})/n + k \log(n)/n$
<b>Hannan-Quinn criterion (HQ)</b>	$-2l(\hat{\beta})/n + 2k \log(\log(n))/n$

where  $l(\hat{\beta})$  is the maximized log likelihood,  $k$  is the number of regressors, and  $n$  is the number of observations. For additional discussion of these and other information criteria, see Grasa (1989) and Appendix F.

## Estimation Options

The iteration limit and convergence criterion may be set in the usual fashion from the Options dialog. In addition, there are options that are specific to binary models. These options are described below.

### Robust standard errors

For binary dependent variable models, EViews allows you to estimate the standard errors using quasi-maximum likelihood (**Huber/White**) or generalized linear model (**GLM**) methods. See the “Technical Notes” on page 427 for a discussion of these two methods.

Click **Options** in the Equation Specification dialog box and check the **Robust Covariance** box and select one of the two methods. When you estimate the binary model using this option, the header in the equation output will indicate the method used to compute the coefficient covariance matrix.

### Starting Values

As with other estimation procedures, EViews allows you to specify starting values. In the options menu, select one of the items from the combo box. You can use the default EViews values, or you can choose a fraction of those values, zero coefficients, or user supplied values. To employ the latter, enter the coefficients in the C coefficient vector, and select **User Supplied** in the combo box.

The EViews default values are selected using a sophisticated algorithm that is specialized for each type of binary model. Unless there is a good reason to choose otherwise, we recommend that you use the default values.

## Estimation Algorithm

By default, EViews uses quadratic hill-climbing to obtain parameter estimates. This algorithm uses the matrix of analytic second derivatives of the log likelihood in forming iteration updates and in computing the estimated covariance matrix of the coefficients.

If you wish, you can employ a different estimation algorithm: Newton-Raphson also employs second derivatives (without the diagonal weighting); BHHH uses first derivatives to determine both iteration updates and the covariance matrix estimates (see Appendix E, page 591 for details). To employ one of these latter methods, click on **Options** in the Equation specification dialog box, and select the desired method.

## Estimation Problems

In general, estimation of binary models is quite straightforward, and you should experience little difficulty in obtaining parameter estimates. There are a few situations, however, where you may experience problems.

First, you may get the error message “Dependent variable has no variance.” This error means that there is no variation in the dependent variable (the variable is always one or zero for all valid observations). This error most often occurs when EViews excludes the entire sample of observations for which  $y$  takes values other than zero or one, leaving too few observations for estimation.

You should make certain to recode your data so that the binary indicators take the values zero and one. This requirement is not as restrictive as it may first seem, since the recoding may easily be done using auto-series. Suppose, for example, that you have data where  $y$  takes the values 1000 and 2000. You could then use the boolean auto-series,  $y=1000$ , or perhaps,  $y<1500$ , as your dependent variable.

Second, you may receive an error message of the form “[xxxx] perfectly predicts binary response [success/failure]”, where xxxx is a sample condition. This error occurs when one of the regressors contains a separating value for which all of the observations with values below the threshold are associated with a single binary response, and all of the values above the threshold are associated with the alternative response. In this circumstance, the method of maximum likelihood breaks down.

For example, if all values of the explanatory variable  $x > 0$  are associated with  $y = 1$ , then  $x$  is a perfect predictor of the dependent variable, and EViews will issue an error message and stop the estimation procedure.

The only solution to this problem is to remove the offending variable from your specification. Usually, the variable has been incorrectly entered in the model, as when a researcher includes a dummy variable that is identical to the dependent variable (for discussion, see Greene, 1997).

Thirdly, you may experience the error, “Non-positive likelihood value observed for observation [xxxx].” This error most commonly arises when the starting values for estimation are poor. The default EViews starting values should be adequate for most uses. You may wish to check the Options dialog to make certain that you are not using user specified starting values, or you may experiment with alternative user-specified values.

Lastly, the error message “Near-singular matrix” indicates that EViews was unable to invert the matrix required for iterative estimation. This will occur if the model is not identified. It may also occur if the current parameters are far from the true values. If you believe the latter to be the case, you may wish to experiment with starting values or the estimation algorithm. The BHHH and quadratic hill-climbing algorithms are less sensitive to this particular problem than is Newton-Raphson.

## Views of Binary Equations

EViews provides a number of standard views and procedures for binary models. For example, you can easily perform Wald or likelihood ratio tests by selecting **View/Coefficient Tests**, and then choosing the appropriate test. In addition, EViews allows you to examine and perform tests using the residuals from your model. The ordinary residuals used in most calculations are described above—additional residual types are defined below. Note that some care should be taken in interpreting test statistics that use these residuals since some of the underlying test assumptions may not be valid in the current setting.

There are a number of additional specialized views and procedures which allow you to examine the properties and performance of your estimated binary model.

## Categorical Regressor Stats

This view displays descriptive statistics (mean and standard deviation) for each regressor. The descriptive statistics are computed for the whole sample, as well as the sample broken down by the value of the dependent variable  $y$ .

## Expectation-Prediction (Classification) Table

This view displays  $2 \times 2$  tables of correct and incorrect classification based on a user specified prediction rule, and on expected value calculations. Click on **View/Expectation-Prediction Table**. EViews opens a dialog prompting you to specify a prediction cutoff value,  $p$ , lying between zero and one. Each observation will be classified as having a predicted probability that lies above or below this cutoff.

After you enter the cutoff value and click on **OK**, EViews will display four (bordered)  $2 \times 2$  tables in the equation window. Each table corresponds to a contingency table of the predicted response classified against the observed dependent variable. The top two tables and associated statistics depict the classification results based given the specified cutoff values. For example:

Dependent Variable: LFP  
 Method: ML - Binary Probit  
 Date: 10/01/97 Time: 23:37  
 Sample: 1 872  
 Included observations: 872

Prediction Evaluation (success cutoff C = 0.5)

	Estimated Equation			Constant Probability		
	Dep=0	Dep=1	Total	Dep=0	Dep=1	Total
P(Dep=1)≤C	321	140	461	471	401	872
P(Dep=1)>C	150	261	411	0	0	0
Total	471	401	872	471	401	872
Correct	321	261	582	471	0	471
% Correct	68.15	65.09	66.74	100.00	0.00	54.01
% Incorrect	31.85	34.91	33.26	0.00	100.00	45.99
Total Gain*	-31.85	65.09	12.73			
Percent Gain**	NA	65.09	27.68			

are the two classification tables based upon a cutoff of 0.5.

In the left-hand table, we classify observations as having predicted probabilities  $\hat{p}_i = 1 - F(-x_i'\hat{\beta})$  that are above or below the specified cutoff value. In the upper right-hand table, we classify observations using  $\bar{p}$ , the sample proportion of  $y = 1$  observations. This probability, which is constant across individuals, is the value computed from estimating a model that includes only the intercept term, C.

“Correct” classifications are obtained when the predicted probability is less than or equal to the cutoff and the observed  $y = 0$ , or when the predicted probability is greater than the cutoff and the observed  $y = 1$ . In the example above, 321 of the Dep = 0 observations and 261 of the Dep = 1 observations are correctly classified by the estimated model.

It is worth noting that in the statistics literature, what we term the expectation-prediction table is sometimes referred to as the *classification table*. The fraction of  $y = 1$  observations that are correctly predicted is termed the *sensitivity*, while the fraction of  $y = 0$  observations that are correctly predicted is known as *specificity*. In EViews, these two values, expressed in percentage terms, are labeled “% Correct”. Overall, the estimated model correctly predicts 66.74 % (68.15 % of the Dep = 0 and 65.09 % of the Dep = 1) observations.

The gain in the number of correct predictions obtained in moving from the right table to the left table provides a measure of the predictive ability of your model. The gain measures are reported in both absolute percentage increases (**Total Gain**), and as a percentage of the incorrect classifications in the constant probability model (**Percent Gain**). In the example above, the restricted model predicts and all 872 individuals will have Dep = 0. This prediction is correct for the 471  $y = 0$  observations, but none of the 401  $y = 1$  observations.



The estimated model improves on the  $\text{Dep} = 1$  predictions by 65.09 percentage points, but does more poorly on the  $\text{Dep} = 0$  predictions (by 31.85 percentage points). Overall, the estimated equation is 12.73 percentage points better at predicting responses than the constant probability model. This represents a 27.68 percent improvement over the 45.09 % correct prediction of the default model.

The bottom portion of the equation window contains analogous prediction results based upon expected value calculations:

	Estimated Equation			Constant Probability		
	Dep=0	Dep=1	Total	Dep=0	Dep=1	Total
E(# of Dep=0)	286.63	183.84	470.46	254.40	216.60	471.00
E(# of Dep=1)	184.37	217.16	401.54	216.60	184.40	401.00
Total	471.00	401.00	872.00	471.00	401.00	872.00
Correct	286.63	217.16	503.79	254.40	184.40	438.81
% Correct	60.85	54.16	57.77	54.01	45.99	50.32
% Incorrect	39.15	45.84	42.23	45.99	54.01	49.68
Total Gain*	6.84	8.17	7.45			
Percent Gain**	14.88	15.12	15.00			

In the left-hand table, we compute the expected number of  $y = 0$  and  $y = 1$  observations in the sample. For example, **E(# of Dep = 0)** is computed as

$$\sum_i \Pr(y_i = 0 | x_i, \beta) = \sum_i F(-x_i' \hat{\beta}), \quad (17.13)$$

where the cumulative distribution function  $F$  is for the normal, logistic, or extreme value distribution.

In the lower right-hand table, we compute the expected number of  $y = 0$  and  $y = 1$  observations for a model estimated with only a constant. For this restricted model, **E(# of Dep = 0)** is computed as  $n(1 - \bar{p})$ , where  $\bar{p}$  is the sample proportion of  $y = 1$  observations. EViews also reports summary measures of the total gain and the percent (of the incorrect expectation) gain.

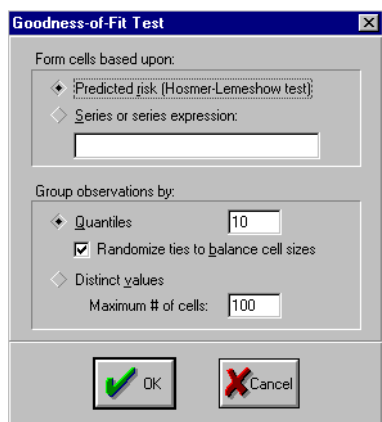
Among the 471 individuals with  $y = 0$ , the expected number of  $y = 0$  observations in the estimated model is 286.63. Among the 401 observations with  $y = 1$ , the expected number of  $y = 1$  observations is 217.16. These numbers represent roughly a seven percentage point (15 percent) improvement over the constant probability model.

### Goodness-of-Fit Tests

This view allows you to perform Pearson  $\chi^2$ -type tests of goodness-of-fit. EViews carries out two goodness-of-fit tests: Hosmer-Lemeshow (1984) and Andrews (1988). The idea underlying these tests is to compare the fitted expected values to the actual values *by group*. If these differences are “large”, we reject the model as providing an insufficient fit to the data.

Details on the two tests are described in the “Technical Notes” on page 427. Briefly, the tests differ in how the observations are grouped and in the asymptotic distribution of the test statistic. The Hosmer-Lemeshow test groups observations on the basis of the predicted probability that  $y = 1$ . The Andrews test is a more general test that groups observations on the basis of any series or series expression.

To carry out the test, select **View/Goodness-of-Fit Test...**



To carry out the goodness-of-fit test, you must first decide on the grouping variable. You can select Hosmer-Lemeshow (predicted probability) grouping by clicking on the corresponding radio button, or you can select series grouping, and provide a series to be used in forming the groups.

Next, you need to specify the grouping rule. EVIEWS allows you to group on the basis of either distinct values or quantiles of the grouping variable.

If your grouping variable takes relatively few distinct values, you should choose the **Distinct values** grouping. EVIEWS will form a separate group for each distinct value of the grouping variable. For example, if your grouping variable is AGE, EVIEWS will create a group for each distinct AGE value and compare the expected

and actual numbers of  $y = 1$  observations in each group. By default, EVIEWS limits you to 100 distinct values. If your grouping series exceeds this value, EVIEWS will return an error message. If you wish to evaluate the test for more than 100 values, you must explicitly increase the maximum number of distinct values.

If your grouping variable takes on a large number of distinct values, you should select **Quantiles**, and enter the number of desired bins in the edit field. If you select this method, EVIEWS will group your observations into the number of specified bins, on the basis of the ordered values of the grouping series. For example, if you group by AGE, select Quantiles, and enter 10, EVIEWS will form groups on the basis of age deciles.

If you choose to group by quantiles and there are ties in the grouping variable, EVIEWS may not be able to form the exact number of groups you specify unless tied values are assigned to different groups. Furthermore, the number of observations in each group may be very unbalanced. Selecting the **randomize ties** option randomly assigns ties to adjacent groups in order to balance the number of observations in each group.

Since the properties of the test statistics require that the number of observations in each group is “large”, some care needs to be taken in selecting a rule so that you do not end up with a large number of cells, each containing small numbers of observations.

By default, EViews will perform the test using Hosmer-Lemeshow grouping. The default grouping method is to form 10 deciles. The test result using the default specification is given by:

Dependent Variable: LFP  
Method: ML - Binary Probit  
Date: 10/01/97 Time: 23:37  
Sample: 1 872  
Included observations: 872  
Andrews and Hosmer-Lemeshow Goodness-of-Fit Tests  
Grouping based upon predicted risk (randomize ties)

	Quantile of Risk		Dep=0		Dep=1		Total Obs	H-L Value
	Low	High	Actual	Expect	Actual	Expect		
1	0.0069	0.1865	74	76.4010	13	10.5990	87	0.61935
2	0.1881	0.2649	64	67.0984	23	19.9016	87	0.62544
3	0.2657	0.3466	63	60.5949	24	26.4051	87	0.31452
4	0.3472	0.4158	60	53.7157	27	33.2843	87	1.92173
5	0.4177	0.4825	48	48.3217	40	39.6783	88	0.00475
6	0.4825	0.5439	41	42.1902	46	44.8098	87	0.06519
7	0.5439	0.5959	40	37.6201	47	49.3799	87	0.26525
8	0.5961	0.6391	35	33.3784	52	53.6216	87	0.12783
9	0.6391	0.6980	29	28.6551	58	58.3449	87	0.00619
10	0.6981	0.9584	17	22.4863	71	65.5137	88	1.79801
Total			471	470.462	401	401.538	872	5.74826
H-L Statistic:			5.74826		Prob[Chi-Sq(8 df)]:		0.67541	
Andrews Statistic:			8.59555		Prob[Chi-Sq(10 df)]:		0.57087	

The columns labeled “Quantiles of Risk” depict the high and low value of the predicted probability for each decile. Also depicted are the actual and expected number of observations in each group, as well as the contribution of each group to the overall Hosmer-Lemeshow (H-L) statistic—large values indicate large differences between the actual and predicted values for that decile.

The  $\chi^2$  statistics are reported at the bottom of the table. Since grouping on the basis of the fitted values falls within the structure of an Andrews test, we report results for both the H-L and the Andrews test statistic. The  $p$ -values for both tests are large and do not indicate significant misspecification of the model.

## Procedures for Binary Equations

In addition to the usual procedures for equations, EViews allows you to forecast the dependent variable and linear index, or to compute a variety of residuals associated with the binary model.

### Forecast

EViews allows you to compute either the fitted probability,  $\hat{p}_i = 1 - F(-x_i'\hat{\beta})$ , or the fitted values of the index  $x_i'\hat{\beta}$ . From the equation toolbar select **Procs/Forecast (Fitted Probability/Index)**..., and then click on the desired entry.

As with other estimators, you can select a forecast sample, and display a graph of the forecast. If your explanatory variables,  $x_t$ , include lagged values of the binary dependent variable  $y_t$ , forecasting with the **Dynamic** option instructs EViews to use the fitted values  $\hat{p}_{t-1}$ , to derive the forecasts, in contrast with the **Static** option, which uses the actual (lagged)  $y_{t-1}$ .

Neither forecast evaluations nor automatic calculation of standard errors of the forecast are currently available for this estimation method. The latter can be computed using the variance matrix of the coefficients displayed by **View/Covariance Matrix**, or using the `@covariance` function.

You can use the fitted index in a variety of ways, for example, to compute the marginal effects of the explanatory variables. Simply forecast the fitted index and save the results in a series, say XB. Then the auto-series `@dnorm(-xb)`, `@dlogistic(-xb)`, or `@dextreme(-xb)` may be multiplied by the coefficients of interest to provide an estimate of the derivatives of the expected value of  $y_i$  with respect to the  $j$ -th variable in  $x_i$ :

$$\frac{\partial E(y_i | x_i, \beta)}{\partial x_{ij}} = f(-x_i' \beta) \beta_j. \quad (17.14)$$

### Make Residual Series

**Procs/Make Residual Series** gives you the option of generating one of the following three types of residuals:

Ordinary	$e_{oi} = y_i - \hat{p}_i$
Standardized	$e_{si} = \frac{y_i - \hat{p}_i}{\sqrt{\hat{p}_i(1 - \hat{p}_i)}}$
Generalized	$e_{gi} = \frac{(y_i - \hat{p}_i)f(-x_i' \hat{\beta})}{\hat{p}_i(1 - \hat{p}_i)}$

where  $\hat{p}_i = 1 - F(-x_i' \hat{\beta})$  is the fitted probability, and the distribution and density functions  $F$  and  $f$ , depend on the specified distribution.

The ordinary residuals have been described above. The standardized residuals are simply the ordinary residuals divided by an estimate of the theoretical standard deviation. The generalized residuals are derived from the first order conditions that define the ML estimates. The first order conditions may be regarded as an orthogonality condition between the generalized residuals and the regressors  $x$ .

$$\frac{\partial l(\beta)}{\partial \beta} = \sum_{i=1}^N \frac{(y_i - (1 - F(-x_i' \beta)))f(-x_i' \beta)}{F(-x_i' \beta)(1 - F(-x_i' \beta))} \cdot x_i = \sum_{i=1}^N e_{gi} \cdot x_i. \quad (17.15)$$

This property is analogous to the orthogonality condition between the (ordinary) residuals and the regressors in linear regression models.

The usefulness of the generalized residuals derives from the fact that you can easily obtain the score vectors by multiplying the generalized residuals by each of the regressors in  $x$ . These scores can be used in a variety of LM specification tests (see Chesher, Lancaster and Irish (1985), and Gourieroux, Monfort, Renault, and Trognon (1987)). We provide an example below.

## Demonstrations

You can easily use the results of a binary model in additional analysis. Here, we provide demonstrations of using EViews to plot a probability response curve and to test for heteroskedasticity in the residuals.

### Plotting probability response curves

You can use the estimated coefficients from a binary model to examine how the predicted probabilities vary with an independent variable. To do so, we will use the EViews built-in modeling features.

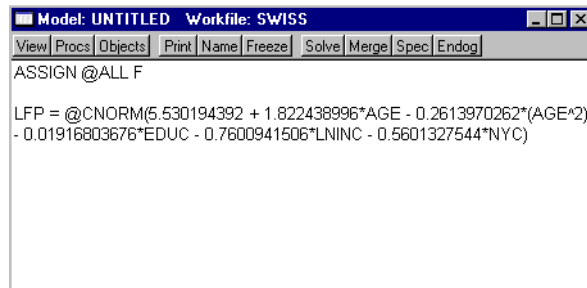
For the probit example above, suppose we are interested in the effect of age (AGE) on labor force participation (LFP). We wish to plot the fitted probabilities as a function of age, fixing the values of other variables at the sample medians.

First, we will create a series containing the values of AGE for which we wish to examine the fitted probabilities. For graphical purposes, we will compute values for AGE that are continuous from 1.8 to 6.5 (ages 18 to 65). The easiest way to do this is to use the `@trend` function:

```
series age_plot=1.8+@trend*(6.5-1.8)/@obs(@trend)
```

`@trend` creates a series that begins at 0 in the first observation of the sample, and increases by 1 for each subsequent observation. Thus, AGE\_PLOT is an equally spaced, increasing series, with values ranging from 1.8 to 6.5.

Next, create a model out of the estimated equation by selecting **Procs/Make Model** from the equation toolbar. EViews creates an untitled model object and opens the model window:



This single equation model describes the formula for generating predicted probabilities based upon the original data and parameter estimates. You can edit this specification to perform the desired simulation by replacing references to EDUC, LNINC and NYC with functions that calculate the median values of those series, and by replacing references to AGE with the constructed series AGE\_PLOT:

```

lfp = @cnorm(5.530194392 + 1.822438996*age_plot -
0.2613970262*(age_plot^2) - 0.01916803676*@median(educ) -
0.7600941506*@median(lninc) - 0.5601327544*@median(nyc))
  
```

Once you have edited your model, click on **Solve** to create the series LFPF containing the forecasted probabilities.

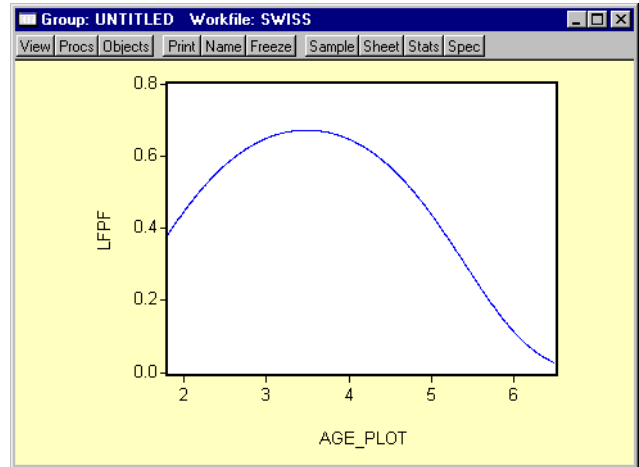
Now you are ready to plot results. Select **Object/New Object/Group**, and enter

```
age_plot lfpf
```

EViews will open an untitled group window containing these two series. Select **View/Graph/XY line** to display the probability of labor force participation plotted against age (with the other variables evaluated at medians).

Here, the quadratic response of predicted labor force participation to age is readily apparent. Participation rates appear to peak at about age 36.

We can also use the model solution features to create more complicated graphs. Suppose that we are interested in examining the effect of the number of young children (NYC) on the relationship between age and labor force participation (LFP). We will need to create a slightly more complicated model to handle this simulation.



First, select **Procs/Make Model** to create a new model. Next, edit the model as before, replacing references to EDUC and LNINC with the functions to compute the median values, and replacing references to AGE with the constructed series AGE\_PLOT. We will create three equations, each representing a distinct value for NYC (you can, of course, use copy-and-paste to aid in creating the additional equations):

```

Model: UNTITLED Workfile: SWISS
View Procs Objects Print Name Freeze Solve Merge Spec Endog

LFP_0 = @CNORM(5.530194392 + 1.822438996*AGE_PLOT -
0.2613970262*(AGE_PLOT^2) - 0.01916803676*@median(EDUC) -
0.7600941506*@median(LNINC) - 0.5601327544*0)

LFP_1 = @CNORM(5.530194392 + 1.822438996*AGE_PLOT -
0.2613970262*(AGE_PLOT^2) - 0.01916803676*@median(EDUC) -
0.7600941506*@median(LNINC) - 0.5601327544*1)

LFP_2 = @CNORM(5.530194392 + 1.822438996*AGE_PLOT -
0.2613970262*(AGE_PLOT^2) - 0.01916803676*@median(EDUC) -
0.7600941506*@median(LNINC) - 0.5601327544*2)

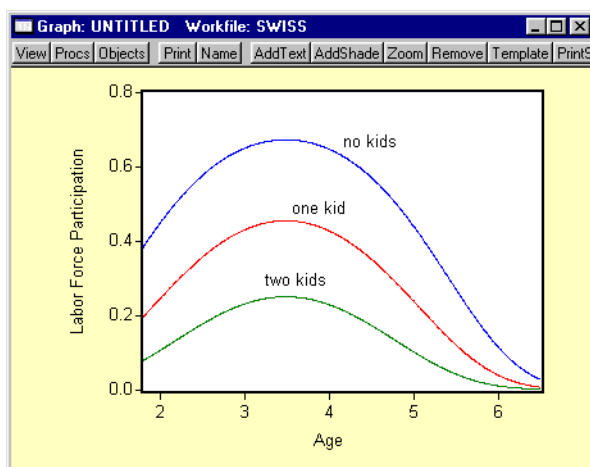
```

Here we have three equations, corresponding to NYC = 0, NYC = 1, and NYC = 2. We have also changed the name of the series to be generated and deleted the ASSIGN statement from the beginning of the file. Click on **Solve** to solve for the values of LFP\_0, LFP\_1, LFP\_2.

To plot these series as an XY-Line graph, select **Object/New Objects/Group**, and enter the names in order

```
age lfp_0 lfp_1 lfp_2
```

EViews will open a group window containing these series. Select **View/Graph/XY line** from the group toolbar to display the following graph:



We have annotated the graph slightly so that you can better judge the effect of number of children on the age-labor force participation relationship.

### Testing for Heteroskedasticity

As an example of specification tests for binary dependent variable models, we carry out the LM test for heteroskedasticity using the artificial regression method described by Davidson and MacKinnon (1993, section 15.4). We test the null hypothesis of homoskedasticity against the alternative of heteroskedasticity of the form

$$\text{var}(u_i) = \exp(2z_i'\gamma), \quad (17.16)$$

where  $\gamma$  is an unknown parameter. In this example, we take the number young children (NYC) as the variable  $z$ . The test statistic is the explained sum of squares from the regression

$$\frac{(y_i - \hat{p}_i)}{\sqrt{\hat{p}_i(1 - \hat{p}_i)}} = \frac{f(-x_i'\hat{\beta})}{\sqrt{\hat{p}_i(1 - \hat{p}_i)}} x_i' b_1 + \frac{f(-x_i'\hat{\beta})(-x_i'\hat{\beta})}{\sqrt{\hat{p}_i(1 - \hat{p}_i)}} z_i' b_2 + v_i, \quad (17.17)$$

which is asymptotically distributed as a  $\chi^2$  with degrees of freedom equal to the number of variables in  $z$  (in this case 1).

To carry out the test, we first retrieve the fitted probabilities  $\hat{p}_i$  and fitted index  $x_i'\hat{\beta}$ . Click on the **Forecast** button and first save the fitted probabilities and then the index as named series in



the workfile (you will have to click **Forecast** twice to save each series). We will assume that the fitted probabilities and index are named P\_HAT and XB, respectively.

Next, the dependent variable in the test regression may be obtained as the standardized residual. Select **Procs/Make Residual Series...** and select **Standardized Residual**. We will save the series as BRMR\_Y.

Lastly, we will use the built-in EViews functions for evaluating the normal density and cumulative distribution function to create a group object containing the independent variables:

```
series fac=@dnorm(-xb)/sqrt(p_hat*(1-p_hat))
group brmr_x fac (age*fac) (educ*fac) (lninc*fac) (nyc*fac)
(nyc*(-xb)*fac)
```

Then run the artificial regression by clicking on **Quick/Estimate Equation...**, selecting **Least Squares**, and entering:

```
brmr_y brmr_x
```

You can obtain the fitted values by clicking on the **Forecast** button in the equation toolbar of this artificial regression. The LM test statistic is the sum of squares of these fitted values. If the fitted values from the artificial regression are saved in BRMR\_YF, the test statistic can be saved as a scalar named LM\_TEST:

```
scalar lm_test=@sumsq(brmr_yf)
```

You can compare the value of this test statistic with the critical values from the chi-square table with one degree of freedom. To save the  $p$ -value as a scalar, enter the command:

```
scalar p_val=1-@cchisq(lm_test,1)
```

To examine the value of P\_VAL, double click on its name in the workfile window; the value will be displayed in the status line at the bottom of the EViews window. The  $p$ -value in this example is essentially zero, so we decisively reject the null hypothesis of homoskedasticity.

## Ordered Dependent Variable Models

EViews estimates the ordered-response model of Aitchison and Silvey (1957) under a variety of assumptions about the latent error distribution. In ordered dependent variable models, the observed  $y$  denotes outcomes representing ordered or ranked categories. For example, we may observe individuals who choose between one of four educational outcomes: less than high school, high school, college, advanced degree. Or we may observe individuals who are employed, partially retired, or fully retired.

As in the binary dependent variable model, we can model the observed response by considering a latent variable  $y_i^*$  that depends linearly on the explanatory variables  $x_i$ :

$$y_i^* = x_i' \beta + \epsilon_i \quad (17.18)$$

where  $\epsilon_i$  are independent and identically distributed random variables. The observed  $y_i$  is determined from  $y_i^*$  using the rule:

$$y_i = \begin{cases} 0 & \text{if } y_i^* \leq \gamma_1 \\ 1 & \text{if } \gamma_1 < y_i^* \leq \gamma_2 \\ 2 & \text{if } \gamma_2 < y_i^* \leq \gamma_3 \\ \vdots & \vdots \\ M & \text{if } \gamma_{M-1} < y_i^* \leq \gamma_M \end{cases} \quad (17.19)$$

It is worth noting that the actual values chosen to represent the categories in  $y$  are completely arbitrary. All the ordered specification requires is for ordering to be preserved so that  $y_i^* < y_j^*$  implies that  $y_i < y_j$ .

It follows that the probabilities of observing each value of  $y$  are given by

$$\begin{aligned} \Pr(y_i = 0 | x_i, \beta, \gamma) &= F(\gamma_1 - x_i' \beta) \\ \Pr(y_i = 1 | x_i, \beta, \gamma) &= F(\gamma_2 - x_i' \beta) - F(\gamma_1 - x_i' \beta) \\ \Pr(y_i = 2 | x_i, \beta, \gamma) &= F(\gamma_3 - x_i' \beta) - F(\gamma_2 - x_i' \beta) \\ &\vdots \\ \Pr(y_i = M | x_i, \beta, \gamma) &= 1 - F(\gamma_M - x_i' \beta) \end{aligned} \quad (17.20)$$

where  $F$  is the cumulative distribution function of  $\epsilon$ .

The threshold values  $\gamma$  are estimated along with the  $\beta$  coefficients by maximizing the log likelihood function:

$$l(\beta, \gamma) = \sum_{i=1}^N \sum_{j=0}^M \log(\Pr(y_i = j | x_i, \beta, \gamma)) \cdot 1(y_i = j) \quad (17.21)$$

where  $1(\cdot)$  is an indicator function which takes the value 1 if the argument is true, and 0 if the argument is false. By default, EViews uses analytic second derivative methods to obtain parameter and variance matrix of the estimated coefficient estimates (see Appendix E, page 591).

## Estimating Ordered Models in EViews

Suppose that the dependent variable DANGER is an index ordered from 1 (least dangerous animal) to 5 (most dangerous animal). We wish to model this ordered dependent variable as a function of the explanatory variables, BODY, BRAIN and SLEEP. Note that the values that we have assigned to the dependent variable are not relevant, only the ordering implied by those values.

EViews will estimate an identical model if the dependent variable is recorded to take the values 10, 234, 3243, 54321, 123456.

To estimate this model, choose **Objects/New Object...** from the main menu and select the **Equation** object from the main menu. From the Equation Specification dialog, select estimation method **ORDERED**. The standard estimation dialog will change to match this specification.

There are three parts to specifying an ordered variable model: the equation specification, the error specification, and the sample specification. First, in the **Equation Specification** field you should type the name of the ordered dependent variable followed by the list of your regressors. In our example, you will enter:

```
danger body brain sleep
```

Ordered estimation only supports specification by list so you may not enter an explicit equation.

Also keep in mind that:

- A separate constant term is not separately identified from the limit points  $\gamma$ , so EViews will ignore any constant term in your specification. Thus, the model

```
danger c body brain sleep
```

- is equivalent to the specification above.
- EViews requires the dependent variable to be integer valued, otherwise you will see an error message, and estimation will stop. This is not, however, a serious restriction, since you can easily convert the series into an integer using `@round`, `@floor` or `@ceil` in an auto-series expression.

Next, select between the ordered logit, ordered probit, and the ordered extreme value models by choosing one of the three distributions for the latent error term.

Lastly, specify the estimation sample.

Now click on **OK**, EViews will estimate the parameters of the model using iterative procedures.

Once the estimation procedure converges, EViews will display the estimation results in the equation window. The first part of the table contains the usual header information, including the assumed error distribution, estimation sample, iteration and convergence information, number of distinct values for  $y$ , and the method of computing the coefficient covariance matrix.

Dependent Variable: DANGER  
 Method: ML - Ordered Probit  
 Date: 09/13/97 Time: 10:00  
 Sample(adjusted): 1 61  
 Included observations: 58  
 Excluded observations: 3 after adjusting endpoints  
 Number of ordered indicator values: 5  
 Convergence achieved after 5 iterations  
 Covariance matrix computed using second derivatives

	Coefficient	Std. Error	z-Statistic	Prob.
BODY	0.006346	0.003262	1.945385	0.0517
BRAIN	-0.003506	0.001822	-1.924244	0.0543
SLEEP	-0.158596	0.040440	-3.921741	0.0001

Below the header information are the coefficient estimates and asymptotic standard errors, and the corresponding  $z$ -statistics and significance levels. The estimated coefficients of the ordered model must be interpreted with care (see Greene (1997, section 19.8) or Johnston and DiNardo (1997, section 13.9)).

The *sign* of  $\hat{\beta}_j$  shows the direction of the change in the probability of falling in the endpoint rankings ( $y = 0$  or  $y = 1$ ) when  $x_{ij}$  changes.  $\Pr(y = 0)$  changes in the *opposite* direction of the sign of  $\hat{\beta}_j$  and  $\Pr(y = M)$  changes in the *same* direction as the sign of  $\hat{\beta}_j$ . The effects on the probability of falling in any of the middle rankings are given by:

$$\frac{\partial \Pr(y = k)}{\partial \beta_j} = \frac{\partial F(\gamma_{k+1} - x_i' \beta)}{\partial \beta_j} - \frac{\partial F(\gamma_k - x_i' \beta)}{\partial \beta_j} \quad (17.22)$$

for  $k = 1, \dots, M-1$ . It is impossible to determine the signs of these terms, *a priori*.

The lower part of the estimation output, labeled “Limit Points”, presents the estimates of the  $\gamma$  coefficients and the associated standard errors and probability values:

LimitPoints				
Limit_2:C(4)	-2.382697	0.512993	-4.644695	0.0000
Limit_3:C(5)	-1.598777	0.484884	-3.297237	0.0010
Limit_4:C(6)	-1.028655	0.465433	-2.210104	0.0271
Limit_5:C(7)	-0.241152	0.445500	-0.541307	0.5883
Akaike info criterion	11.74480	Schwarz criterion	11.99348	
Log likelihood	-333.5993	Hannan-Quinn criter.	11.84167	
Avg. log likelihood	-5.751712			

Note that the coefficients are labeled both with the identity of the limit point, and the coefficient number. Just below the limit points are the summary statistics for the equation.

## Estimation Problems

Most of the previous discussion of estimation problems for binary models (p. Estimation Problems) also holds for ordered models. In particular, the models are generally well-behaved and will require little intervention.

There are cases, however, where problems will arise. First, EViews currently has a limit of 750 total coefficients in an ordered dependent variable model. Thus, if you have 25 right-hand side variables, and a dependent variable with 726 distinct values, you will be unable to estimate your model using EViews.

Second, you may run into identification problems and estimation difficulties if you have some groups where there are very few observations. If necessary, you may choose to combine adjacent groups and re-estimate the model.

EViews may stop estimation with the message “Parameter estimates for limit points are non-ascending”, most likely on the first iteration. This error indicates that parameter values for the limit points were invalid, and that EViews was unable to adjust these values to make them valid. Make certain that if you are using user defined parameters, the limit points are strictly increasing. Better yet, we recommend that you employ the EViews starting values since they are based on a consistent first-stage estimation procedure, and should therefore be quite well-behaved.

## Views of Ordered Equations

EViews provides you with several views of an ordered equation. As with other equations, you can examine the specification and estimated covariance matrix as well as perform Wald and likelihood ratio tests on coefficients of the model. In addition, there are several views that are specialized for the ordered model:

**Dependent Variable Frequencies** — computes a one-way frequency table for the ordered dependent variable for the observations in the estimation sample. EViews presents both the frequency table and the cumulative frequency table in levels and percentages.

**Expectation-Prediction Table** — classifies observations on the basis of the predicted response. EViews performs the classification on the basis of maximum predicted probability as well as the expected probability.

Dependent Variable: DANGER  
 Method: ML - Ordered Probit  
 Date: 09/13/97 Time: 10:00  
 Sample(adjusted): 1 61  
 Included observations: 58  
 Excluded observations: 3 after adjusting endpoints  
 Prediction table for ordered dependent variable

Value	Count	Count of obs with Max Prob	Error	Sum of all Probabilities	Error
1	18	27	-9	18.571	-0.571
2	14	16	-2	13.417	0.583
3	10	0	10	9.163	0.837
4	9	8	1	8.940	0.060
5	7	7	0	7.909	-0.909

There are two columns labeled “Error”. The first measures the difference between the observed count and the number of observations where the probability of that response is highest. For example, 18 individuals reported a value of 1 for DANGER, while 27 individuals had predicted probabilities that were highest for this value. The actual count minus the predicted is  $-9$ . The second error column measures the difference between the actual number of individuals reporting the value, and the sum of all of the individual probabilities for that value.

## Procedures for Ordered Equations

### Make ordered limit vector/matrix

The full set of coefficients and the covariance matrix may be obtained from the estimated equation in the usual fashion (see Chapter 11). In some circumstances, however, you may wish to perform inference using only the estimates of the  $\gamma$  coefficients and the associated covariances.

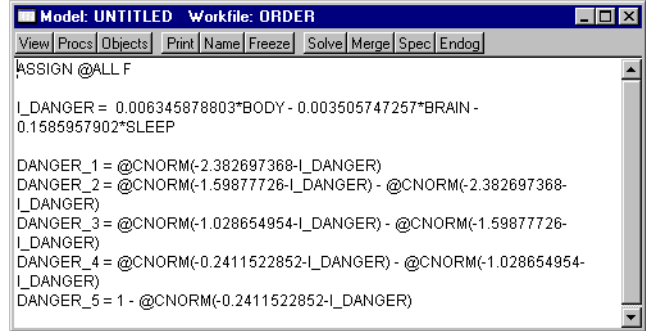
The **Make Ordered Limit Vector** and **Make Ordered Limit Covariance Matrix** procedures provide a shortcut method of obtaining the estimates associated with the  $\gamma$  coefficients. The first procedure creates a vector (using the next unused name of the form LIMITS01, LIMITS02, etc.) containing the estimated  $\gamma$  coefficients. The latter procedure creates a symmetric matrix containing the estimated covariance matrix of the  $\gamma$ . The matrix will be given an unused name of the form VLIMITS01, VLIMITS02, etc., where the “V” is used to indicate that these are the variances of the estimated limit points.

### Forecasting using Models

You cannot forecast directly from an estimated ordered model since the dependent variable represents categorical or rank data. EViews does, however, allow you to forecast the probability associated with each category. To forecast these probabilities, you must first create a model.

Choose **Procs/Make Model** and EViews will open an untitled model window containing a system of equations, with a separate equation for the probability of each ordered response value:

The first line of the model is the assignment statement which indicates that the forecasted series will have a name with a letter F appended to each left-hand side series name. See Chapter 23 for additional detail on modifying the assignment statement and additional information on solving models.



To forecast from this model, simply click the **Solve** button in the model window toolbar.

For this example, the series  $I\_DANGERF$  will contain the fitted linear index  $x_i'\hat{\beta}$ . The fitted probability of falling in category 1 will be stored as a series named  $DANGER\_1F$ , the fitted probability of falling in category 2 will be stored as a series named  $DANGER\_2F$ , and so on. Note that for each observation, the fitted probability of falling in each of the categories sums up to one.

### Make Residual Series

The generalized residuals of the ordered model are the derivatives of the log likelihood with respect to a hypothetical unit- $x$  variable. These residuals are defined to be uncorrelated with the explanatory variables of the model (see Chesher and Irish (1987), and Gourieroux, Monfort, Renault and Trognon (1987) for details), and thus may be used in a variety of specification tests.

To create a series containing the generalized residuals, select **View/Make Residual Series...**, enter a name or accept the default name, and click **OK**. The generalized residuals for an ordered model are given by:

$$e_{gi} = \frac{f(\gamma_g - x_i'\hat{\beta}) - f(\gamma_{g-1} - x_i'\hat{\beta})}{F(\gamma_g - x_i'\hat{\beta}) - F(\gamma_{g-1} - x_i'\hat{\beta})}, \quad (17.23)$$

where  $\gamma_0 = -\infty$ , and  $\gamma_{M+1} = \infty$ .

## Censored Regression Models

In some settings, the dependent variable is only partially observed. For example, in survey data, data on incomes above a specified level are often top-coded to protect confidentiality. Similarly desired consumption on durable goods may be censored at a small positive or zero value.

EViews provides tools to perform maximum likelihood estimation of these models and to use the results for further analysis.

## Theory

Consider the following latent variable regression model

$$y_i^* = x_i' \beta + \sigma \epsilon_i, \quad (17.24)$$

where  $\sigma$  is a scale parameter. The scale parameter  $\sigma$  is identified in censored and truncated regression models, and will be estimated along with the  $\beta$ .

In the canonical *censored regression model*, known as the *tobit*, the observed data  $y$  are given by

$$y_i = \begin{cases} 0 & \text{if } y_i^* \leq 0 \\ y_i^* & \text{if } y_i^* > 0 \end{cases} \quad (17.25)$$

In other words, all negative values of  $y_i^*$  are coded as 0. We say that these data are *left censored*.

at 0. Note that this situation differs from a *truncated regression model* where negative values of  $y_i^*$  are dropped from the sample. More generally, EViews allows for both left and right censoring at arbitrary limit points so that

$$y_i = \begin{cases} \underline{c}_i & \text{if } y_i^* \leq \underline{c}_i \\ y_i^* & \text{if } \underline{c}_i < y_i^* \leq \bar{c}_i \\ \bar{c}_i & \text{if } \bar{c}_i < y_i^* \end{cases} \quad (17.26)$$

where  $\underline{c}_i$ ,  $\bar{c}_i$  are fixed numbers representing the censoring points. If there is no left censoring, then we can set  $\underline{c}_i = -\infty$ . If there is no right censoring, then  $\bar{c}_i = \infty$ . The canonical tobit model is a special case with  $\underline{c}_i = 0$  and  $\bar{c}_i = \infty$ .

The parameters  $\beta$ ,  $\sigma$  are estimated by maximizing the log likelihood function

$$\begin{aligned} l(\beta, \sigma) = & \sum_{i=1}^N \log f((y_i - x_i' \beta) / \sigma) \cdot 1(\underline{c}_i < y_i < \bar{c}_i) + \log(F((\underline{c}_i - x_i' \beta) / \sigma)) \cdot 1(y_i = \underline{c}_i) \\ & + \log(1 - F((\bar{c}_i - x_i' \beta) / \sigma)) \cdot 1(y_i = \bar{c}_i) \end{aligned} \quad (17.27)$$

where  $f$ ,  $F$  are the density and cumulative distribution functions of  $\epsilon$ .

## Estimating Censored Models in EViews

Consider the model



$$\text{HRS}_i = \beta_1 + \beta_2 \text{AGE}_i + \beta_3 \text{EDU}_i + \beta_4 \text{KID}_i + \epsilon_i, \quad (17.28)$$

where hours worked (HRS) is left censored at zero. To estimate this model, select **Quick/Estimate Equation...** from the main menu. Then from the Equation Specification dialog, select the **CENSORED** estimation method. The dialog will change to provide a number of different input options.

### Specifying the Regression Equation

In the **Equation Specification** field, enter the name of the censored dependent variable followed by a list of regressors. In our example, you will enter

hrs c age edu kid

Censored estimation only supports specification by list so you may not enter an explicit equation.

Next, select one of the three distributions for the error term. EViews allows you three possible choices for the distribution of  $\epsilon$ :

Standard normal	$E(\epsilon) = 0, \text{var}(\epsilon) = 1$
Logistic	$E(\epsilon) = 0, \text{var}(\epsilon) = \pi^2/3$
Extreme value (Type I)	$E(\epsilon) \approx 0.5772$ (Euler's constant), $\text{var}(\epsilon) = \pi^2/6$

### Specifying the censoring points

You must also provide information about the censoring points of the dependent variable. There are two cases to consider: (1) where the limit points are known for all individuals, and (2) where the censoring is by indicator and the limit points are known only for individuals with censored observations.

#### Limit Points Known

You should enter an expressions for the left and right censoring points in the edit fields as required. Note that if you leave an edit field blank, EViews will assume that there is no censoring of observations of that type.

For example, in the canonical tobit model the data are censored on the left at zero, and are uncensored on the right. This case may be specified as:

Left edit field: 0

Right edit field: [blank]

Similarly, top-coded censored data may be specified as:

Left edit field: [blank]

Right edit field: 20000

while the more general case of left and right censoring is given by:

Left edit field: 10000

Right edit field: 20000

EViews also allows more general specifications where the censoring points are known but differ across observations. Simply enter the name of the series or auto-series containing the censoring points in the appropriate edit field. For example,

Left edit field: lowinc

Right edit field: vcens1+10

specifies a model with LOWINC censoring on the left-hand side, and right censoring at the value of VCENS1 + 10.

### Limit Points Not Known

In some cases, the hypothetical censoring point is unknown for some individuals ( $c_i$  and  $\bar{c}_i$  are not observed for all observations). This situation often occurs with data where censoring is indicated with a zero-one dummy variable, but no additional information is provided about potential censoring points.

EViews provides you an alternative method of describing data censoring that matches this format. Simply select the **Field is zero/one indicator of censoring** option in the estimation dialog, and enter the series expression for the censoring indicator(s) in the appropriate edit field(s). Observations with a censoring indicator of one are assumed to be censored while those with a value of zero are assumed to be actual responses.

For example, suppose that we have observations on the length of time that an individual has been unemployed (U), but that some of these observations represent ongoing unemployment at the time the sample is taken. These latter observations may be treated as right censored at the reported value. If the variable RCENS is a dummy variable representing censoring, you can click on the **Field is zero/one indicator of censoring** setting and enter:

Left edit field: [blank]

Right edit field: rcens

in the edit fields. If the data are censored on both the left and the right, use separate binary indicators for each form of censoring:

Left edit field: lcens

Right edit field: rcens

where LCENS is also a binary indicator.

Once you have specified the model, click **OK**. EViews will estimate the parameters of the model using appropriate iterative techniques.

### A Comparison of Censoring Methods

An alternative to specifying index censoring is to enter a very large positive or negative value for the censoring limit for non-censored observations. For example, you could enter  $-1e100$  and  $1e100$  as the censoring limit for an observation on a completed unemployment spell. In fact, any limit point that is “outside” the observed data will suffice.

While this latter approach will yield the same likelihood function and therefore the same parameter values and coefficient covariance matrix, there is a drawback to the artificial data approach. The presence of a censoring value implies you can evaluate the conditional mean of the observed dependent variable, as well as the ordinary and standardized residuals. All of the calculations that use residuals will, however, be based upon the arbitrary artificial data and will therefore be invalid.

If you specify your censoring by index, you are telling EViews that you do not have information about the censoring for those observations that are not censored. And if an observation is left censored, you may not have information about the right censoring limit. In these circumstances, you should specify your censoring by index so that EViews will prevent you from computing the conditional mean of the dependent variable and the associated residuals.

### Interpreting the Output

If your model converges, EViews will display the estimation results in the equation window. The first part of the table presents the usual header information, including information about the assumed error distribution, estimation sample, estimation algorithms, and number of iterations required for convergence.

EViews also provides information about the specification for the censoring. If the estimated model is the canonical tobit with left-censoring at zero, EViews will label the method as a TOBIT. For all other censoring methods, EViews will display detailed information about form of the left and/or right censoring.

Here, we have the header output from a left censored model where the censoring is specified by value:

```
Dependent Variable: Y_PT
Method: ML - Censored Normal (TOBIT)
Date: 10/13/97   Time: 22:42
Sample: 1 601
Included observations: 601
Left censoring (value) at zero
Convergence achieved after 9 iterations
Covariance matrix computed using second derivatives
```

Below the header are the usual results for the coefficients, including the asymptotic standard errors,  $z$ -statistics, and significance levels. As in other limited dependent variable models, the estimated coefficients do not have a direct interpretation as the marginal effect of the associated regressor  $j$  for individual  $i$ ,  $x_{ij}$ . In censored regression models, a change in  $x_{ij}$  has two effects: an effect on the mean of  $y$ , given that it is observed, and an effect on the probability of  $y$  being observed (see McDonald and Moffitt, 1980).

In addition to results for the regression coefficients, EViews reports an additional coefficient named SCALE, which is the estimated scale factor  $\sigma$ . This scale factor may be used to estimate the standard deviation of the residual, using the known variance of the assumed distribution. For example, if the estimated SCALE has a value of .446 for a model with extreme value errors, the implied standard error of the error term is  $0.466 \times \pi / \sqrt{6} \cong 0.572$ .

Most of the other output is self-explanatory. As in the binary and ordered models above, EViews reports summary statistics for the dependent variable and likelihood based statistics. The regression statistics at the bottom of the table are computed in the usual fashion, using the residuals  $\hat{\epsilon}_i = y_i - E(y_i|x_i, \hat{\beta}, \hat{\sigma})$  from the observed  $y$ .

Views of Censored Equations

Most of the views that are available for a censored regression are familiar from other settings. The residuals used in the calculations are defined below.

The one new view is the **Categorical Regressor Stats** view, which presents means and standard deviations for the dependent and independent variables for the estimation sample. EViews provides statistics computed over the entire sample, as well as for the left censored, right censored and non-censored individuals.

Procedures for Censored Equations

EViews provides several procedures which provide access to information derived from your censored equation estimates.

Make Residual Series

Select **Procs/Make Residual Series**, and select from among the three types of residuals. The three types of residuals for censored models are defined as:

Ordinary	$e_{oi} = y_i - E(y_i x_i, \hat{\beta}, \hat{\sigma})$
----------	--

Standardized	$e_{si} = \frac{y_i - E(y_i x_i, \hat{\beta}, \hat{\sigma})}{\sqrt{\text{var}(y_i x_i, \hat{\beta}, \hat{\sigma})}}$
Generalized	$e_{gi} = -\frac{f((\underline{c}_i - x_i'\hat{\beta})/\hat{\sigma})}{\sigma F((\underline{c}_i - x_i'\hat{\beta})/\hat{\sigma})} \cdot 1(y_i \leq \underline{c}_i) \\ -\frac{f'((\underline{c}_i - x_i'\hat{\beta})/\hat{\sigma})}{\sigma F((\underline{c}_i - x_i'\hat{\beta})/\hat{\sigma})} \cdot 1(\underline{c}_i < y_i \leq \bar{c}_i) \\ +\frac{f((\bar{c}_i - x_i'\hat{\beta})/\hat{\sigma})}{\sigma(1-F((\bar{c}_i - x_i'\hat{\beta})/\hat{\sigma}))} \cdot 1(y_i \leq \bar{c}_i)$

where  $f$ ,  $F$  are the density and distribution functions, and where 1 is an indicator function which takes the value 1 if the condition in parentheses is true, and 0 if it is false. All of the above terms will be evaluated at the estimated  $\beta$  and  $\sigma$ . See the discussion of forecasting, that follows, for details on the computation of  $E(y_i|x_i, \beta, \sigma)$ .

The generalized residuals may be used as the basis of a number of LM tests, including LM tests of normality (see Lancaster, Chesher and Irish (1985), Chesher and Irish (1987), and Gourioux, Monfort, Renault and Trognon (1987); Greene (1997), provides a brief discussion and additional references).

## Forecasting

EViews provides you with the option of forecasting the expected dependent variable,  $E(y_i|x_i, \beta, \sigma)$ , or the expected latent variable,  $E(y_i^*|x_i, \beta, \sigma)$ . Select **Forecast** from the equation toolbar to open the forecast dialog.

To forecast the expected *latent variable*, click on **Index - Expected latent variable**, and enter a name for the series to hold the output. The forecasts of the expected latent variable  $E(y_i^*|x_i, \beta, \sigma)$  may be derived from the latent model using the relationship

$$\hat{y}_i^* = E(y_i^*|x_i, \hat{\beta}, \hat{\sigma}) = x_i'\hat{\beta}. \quad (17.29)$$

To forecast the expected *observed dependent variable*, you should select **Expected dependent variable**, and enter a series name. These forecasts are computed using the relationship:

$$\hat{y}_i = E(y_i|x_i, \hat{\beta}, \hat{\sigma}) = \underline{c}_i \cdot \Pr(y_i = \underline{c}_i|x_i, \hat{\beta}, \hat{\sigma}) \\ + E(y_i^*|\underline{c}_i < y_i^* < \bar{c}_i; x_i, \hat{\beta}, \hat{\sigma}) \cdot \Pr(\underline{c}_i < y_i^* < \bar{c}_i|x_i, \hat{\beta}, \hat{\sigma}) + \bar{c}_i \cdot \Pr(y_i = \bar{c}_i|x_i, \hat{\beta}, \hat{\sigma}) \quad (17.30)$$

Note that these forecasts always satisfy  $\underline{c}_i \leq \hat{y}_i \leq \bar{c}_i$ . The probabilities associated with being in the various classifications are computed by evaluating the cumulative distribution function of the specified distribution. For example, the probability of being at the lower limit is given by:

$$\Pr(y_i = c_i | x_i, \hat{\beta}, \hat{\sigma}) = \Pr(y_i^* \leq c_i | x_i, \hat{\beta}, \hat{\sigma}) = F((c_i - x_i' \hat{\beta}) / \hat{\sigma}). \quad (17.31)$$

### Censored Model Illustration

As an example, we replicate Fair's (1978) tobit model that estimates the incidence of extramarital affairs. The dependent variable, number of extramarital affairs (Y\_PT), is left censored at zero and the errors are assumed to be normally distributed. The bottom portion of the output is presented below:

	Coefficient	Std. Error	z-Statistic	Prob.
C	7.608487	3.905837	1.947979	0.0514
Z1	0.945787	1.062824	0.889881	0.3735
Z2	-0.192698	0.080965	-2.380015	0.0173
Z3	0.533190	0.146602	3.636997	0.0003
Z4	1.019182	1.279524	0.796532	0.4257
Z5	-1.699000	0.405467	-4.190231	0.0000
Z6	0.025361	0.227658	0.111399	0.9113
Z7	0.212983	0.321145	0.663198	0.5072
Z8	-2.273284	0.415389	-5.472657	0.0000
Error Distribution				
SCALE:C(10)	8.258432	0.554534	14.89256	0.0000
R-squared	0.151569	Mean dependent var	1.455907	
Adjusted R-squared	0.138649	S.D. dependent var	3.298758	
S.E. of regression	3.061544	Akaike info criterion	2.378473	
Sum squared resid	5539.472	Schwarz criterion	2.451661	
Log likelihood	-704.7311	Hannan-Quinn criter.	2.406961	
Avg. log likelihood	-1.172597			
Left censored obs	451	Right censored obs	0	
Uncensored obs	150	Total obs	601	

### Tests of significance

EViews does not, by default, provide you with the usual likelihood ratio test of the overall significance for the tobit and other censored regression models. There are several ways to perform this test (or an asymptotically equivalent test).

First, you can use the built-in coefficient testing procedures to test the exclusion of all of the explanatory variables. Select **View/Coefficient Tests/Redundant Variables - Likelihood ratio**, and enter the names of all of the explanatory variables you wish to exclude. EViews will compute the appropriate likelihood ratio test statistic and the *p*-value associated with the statistic.

To take an example, suppose we wish to test whether the variables in the Fair tobit, above, contribute to the fit of the model. Select **View/Coefficient Tests/Redundant Variables - Likelihood Ratio...**, and enter all of the explanatory variables:

z1 z2 z3 z4 z5 z6 z7 z8

EViews will estimate the restricted model for you and compute the LR statistic and  $p$ -value. In this case, the value of the test statistic is 80.01, which for eight degrees-of-freedom, yields a  $p$ -value of less than .000001.

Alternatively, you could test the restriction using the Wald test by selecting **View/Coefficient Tests/Wald - Coefficient Restrictions...**, and entering the restriction that:

$$c(2) = c(3) = c(4) = c(5) = c(6) = c(7) = c(8) = c(9) = 0$$

The reported statistic is 68.14, with a  $p$ -value of less than .000001.

Lastly, we demonstrate the direct computation of the LR test. Suppose the Fair tobit model estimated above is saved in the named equation EQ\_TOBIT. Then you could estimate an equation containing only a constant, say EQ\_RESTR, and place the likelihood ratio statistic in a scalar:

```
scalar lrstat=-2*(eq_restr.@logl-eq_tobit.@logl)
```

Next, evaluate the chi-square probability associated with this statistic:

```
scalar lrprob=1-@cchisq(lrstat, 8)
```

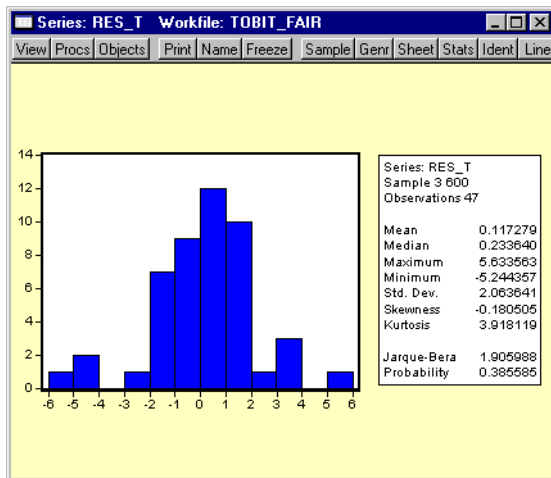
with degrees of freedom given by the number of coefficient restrictions in the constant only model. You can double click on the LRSTAT icon or the LRPROB icon in the workfile window to display the results in the status line.

### A specification test for the Tobit

As a rough diagnostic check, Pagan and Vella (1989) suggest plotting Powell's (1986) symmetrically trimmed residuals. If the error terms have a symmetric distribution centered at zero (as assumed by the normal distribution), so should the trimmed residuals. To construct the trimmed residuals, first save the forecasts of the index (expected latent variable): click **Forecast**, choose **Index-Expected latent variable**, and provide a name for the fitted index, say XB. The trimmed residuals are obtained by dropping observations for which  $x_i'\hat{\beta} < 0$ , and replacing  $y_i$  with  $2(x_i'\hat{\beta})$  for all observations where  $y_i < 2(x_i'\hat{\beta})$ . The trimmed residuals RES\_T can be obtained by the commands

```
series res_t=(y_pt<=2*xb)*(y_pt-xb) +(y_pt>2*xb)*xb
smpl if xb<0
series res_t=na
smpl @all
```

The histogram of the trimmed residual is depicted below.



This example illustrates the possibility that the number of observations that are lost by trimming can be quite large; out of the 601 observations in the sample, only 47 observations are left after trimming.

The tobit model imposes the restriction that the coefficients that determine the probability of being censored are the same as those that determine the conditional mean of the uncensored observations. To test this restriction, we carry out the LR test by comparing the (restricted) tobit to the unrestricted log likelihood that is the sum of a probit and a truncated regression (we discuss truncated regression in detail in the following section). Save the tobit equation in the

workfile by pressing the **Name** button, and enter a name, say EQ\_TOBIT.

To estimate the probit, first create a dummy variable indicating uncensored observations by the command

```
series y_c=(y_pt>0)
```

Then estimate a probit by replacing the dependent variable Y\_PT by Y\_C. A simple way to do this is to press **Objects/Copy Object...** from the tobit equation toolbar. From the new untitled equation window that appears, press **Estimate**, replace the dependent variable with Y\_C and choose **Method: BINARY** and click **OK**. Save the probit equation by pressing the **Name** button, say as EQ\_BIN.

To estimate the truncated model, press **Objects/Copy Object...** from the tobit equation toolbar again. From the new untitled equation window that appears, press **Estimate**, mark the **Truncated sample** option, and click **OK**. Save the truncated regression by pressing the **Name** button, say as EQ\_TR.

Then the LR test statistic and its *p*-value can be saved as a scalar by the commands

```
scalar lr_test=2*(eq_bin.@logl+eq_tr.@logl-eq_tobit.@logl)
scalar lr_pval=1-@cchisq(lr_test,eq_tobit.@ncoef)
```

Double click on the scalar name to display the value in the status line at the bottom of the EViews window. For the example data set, the *p*-value is 0.066 which rejects the tobit model at the 10 % level but not at the 5 % level.

For other specification tests for the tobit, see Greene (1997, 20.3.4) or Pagan and Vella (1989).



## Truncated Regression Models

A close relative of the censored regression model is the truncated regression model. Suppose that an observation is not observed whenever the dependent variable falls below one threshold, or exceeds a second threshold. This sampling rule occurs, for example, in earnings function studies for low-income families that exclude observations with incomes above a threshold, and in studies of durables demand among individuals who purchase durables.

The general two-limit truncated regression model may be written as:

$$y_i^* = x_i' \beta + \epsilon_i \quad (17.32)$$

where  $y_i = y_i^*$  is only observed if:

$$c_i < y_i^* < \bar{c}_i. \quad (17.33)$$

If there is no lower truncation, then we can set  $c_i = -\infty$ . If there is no upper truncation, then we set  $\bar{c}_i = \infty$ .

The log likelihood function associated with these data is given by

$$l(\beta, \sigma) = \sum_{i=1}^N \log f((y_i - x_i' \beta) / \sigma) \cdot 1(c_i < y_i < \bar{c}_i) - \log(F((\bar{c}_i - x_i' \beta) / \sigma) - F((c_i - x_i' \beta) / \sigma)) \quad (17.34)$$

The likelihood function is maximized with respect to  $\beta$  and  $\sigma$ , using standard iterative methods.

### Estimating a truncated model in EViews

Estimation of a truncated regression model follows the same steps as estimating a censored regression:

- Select **Quick/Estimate Equation...** from the main menu, and in the Equation Specification dialog, select the **CENSORED** estimation method. The censored and truncated regression dialog will appear.
- Enter the name of the truncated dependent variable and the list of the regressors in the **Equation Specification** field, and select one of the three distributions for the error term. You must enter your specification by list.
- Indicate that you wish to estimate the truncated model by checking the **Truncated sample** option.
- Specify the truncation points of the dependent variable by entering the appropriate expressions in the two edit fields. If you leave an edit field blank, EViews will assume that there is no truncation along that dimension.

You should keep a few points in mind. First, truncated estimation is only available for models where the truncation points are known, since the likelihood function is not otherwise defined. If you attempt to specify your truncation points by index, EViews will issue an error message indicating that this selection is not available.

Second, EViews will issue an error message if any values of the dependent variable are outside the truncation points. Furthermore, EViews will automatically exclude any observations that are exactly equal to a truncation point. Thus, if you specify zero as the lower truncation limit, EViews will issue an error message if any observations are less than zero, and will exclude any observations where the dependent variable exactly equals zero.

The cumulative distribution function and density of the assumed distribution will be used to form the likelihood function, as described above.

## Procedures for Truncated Equations

EViews provides the same procedures for truncated equations as for censored equations. The residual and forecast calculations differ to reflect the truncated dependent variable and the different likelihood function.

### Make Residual Series

Select **Procs/Make Residual Series**, and select from among the three types of residuals. The three types of residuals for censored models are defined as:

Ordinary	$e_{oi} = y_i - E(y_i^*   \underline{c}_i < y_i^* < \bar{c}_i; x_i, \hat{\beta}, \hat{\sigma})$
Standardized	$e_{si} = \frac{y_i - E(y_i^*   \underline{c}_i < y_i^* < \bar{c}_i; x_i, \hat{\beta}, \hat{\sigma})}{\sqrt{\text{var}(y_i^*   \underline{c}_i < y_i^* < \bar{c}_i; x_i, \hat{\beta}, \hat{\sigma})}}$
Generalized	$e_{gi} = - \frac{f'((y_i - x_i' \hat{\beta}) / \hat{\sigma})}{\sigma f((y_i - x_i' \hat{\beta}) / \hat{\sigma})} - \frac{f((\bar{c}_i - x_i' \hat{\beta}) / \sigma) - f((\underline{c}_i - x_i' \hat{\beta}) / \sigma)}{\sigma (F((\bar{c}_i - x_i' \hat{\beta}) / \sigma) - F((\underline{c}_i - x_i' \hat{\beta}) / \sigma))}$

where  $f$ ,  $F$  are the density and distribution functions. Details on the computation of  $E(y_i | \underline{c}_i < y_i < \bar{c}_i; x_i, \hat{\beta}, \hat{\sigma})$  are provided below.

The generalized residuals may be used as the basis of a number of LM tests, including LM tests of normality (see Chesher and Irish (1984, 1987), and Gouriéroux, Monfort and Trognon (1987); Greene (1997) provides a brief discussion and additional references).

## Forecasting

EViews provides you with the option of forecasting the expected observed dependent variable,  $E(y_i|x_i, \hat{\beta}, \hat{\sigma})$ , or the expected latent variable,  $E(y_i^*|x_i, \hat{\beta}, \hat{\sigma})$ .

To forecast the expected latent variable, select **Forecast** from the equation toolbar to open the forecast dialog, click on **Index - Expected latent variable**, and enter a name for the series to hold the output. The forecasts of the expected latent variable  $E(y_i^*|x_i, \hat{\beta}, \hat{\sigma})$  are computed using

$$\hat{y}_i^* = E(y_i^*|x_i, \hat{\beta}, \hat{\sigma}) = x_i' \hat{\beta}. \quad (17.35)$$

To forecast the expected observed dependent variable for the truncated model, you should select **Expected dependent variable**, and enter a series name. These forecasts are computed using:

$$\hat{y}_i = E(y_i^*|c_i < y_i^* < \bar{c}_i; x_i, \hat{\beta}, \hat{\sigma}) \quad (17.36)$$

so that the expectations for the latent variable are taken with respect to the conditional (on being observed) distribution of the  $y_i^*$ . Note that these forecasts always satisfy  $c_i < \hat{y}_i < \bar{c}_i$ .

It is instructive to compare this latter expected value with the expected value derived for the censored model in Equation (17.30) above (repeated here for convenience):

$$\begin{aligned} \hat{y}_i = E(y_i|x_i, \hat{\beta}, \hat{\sigma}) &= c_i \cdot \Pr(y_i = c_i|x_i, \hat{\beta}, \hat{\sigma}) \\ &+ E(y_i^*|c_i < y_i^* < \bar{c}_i; x_i, \hat{\beta}, \hat{\sigma}) \cdot \Pr(c_i < y_i^* < \bar{c}_i|x_i, \hat{\beta}, \hat{\sigma}) + \bar{c}_i \cdot \Pr(y_i = \bar{c}_i|x_i, \hat{\beta}, \hat{\sigma}) \end{aligned} \quad (17.37)$$

The expected value of the dependent variable for the truncated model is the first part of the middle term of the censored expected value. The differences between the two expected values (the probability weight and the first and third terms) reflect the different treatment of latent observations that do not lie between  $c_i$  and  $\bar{c}_i$ . In the censored case, those observations are included in the sample and are accounted for in the expected value. In the truncated case, data outside the interval are not observed and are not used in the expected value computation.

## Illustration

As an example, we reestimate the Fair tobit model from above, truncating the data so that observations at or below zero are removed from the sample. The output from truncated estimation of the Fair model is presented below:

Dependent Variable: Y_PT				
Method: ML - Censored Normal (TOBIT)				
Date: 10/13/97 Time: 22:45				
Sample(adjusted): 452 601				
Included observations: 150 after adjusting endpoints				
Truncated sample				
Left censoring (value) at zero				
Convergence achieved after 8 iterations				
Covariance matrix computed using second derivatives				
	Coefficient	Std. Error	z-Statistic	Prob.
C	12.37288	5.178306	2.389368	0.0169
Z1	-1.336872	1.453133	-0.919993	0.3576
Z2	-0.044792	0.116141	-0.385670	0.6997
Z3	0.544182	0.220119	2.472218	0.0134
Z4	-2.142896	1.787720	-1.198675	0.2307
Z5	-1.423128	0.600472	-2.370014	0.0178
Z6	-0.316721	0.322327	-0.982609	0.3258
Z7	0.621428	0.478827	1.297813	0.1944
Z8	-1.210037	0.552131	-2.191578	0.0284
Error Distribution				
SCALE:C(10)	5.379557	0.688875	7.809196	0.0000
R-squared	0.654664	Mean dependent var	1.455907	
Adjusted R-squared	0.649405	S.D. dependent var	3.298758	
S.E. of regression	1.953229	Akaike info criterion	1.333891	
Sum squared resid	2254.726	Schwarz criterion	1.407079	
Log likelihood	-390.8342	Hannan-Quinn criter.	1.362379	
Avg. log likelihood	-0.650306			
Left censored obs	0	Right censored obs	0	
Uncensored obs	150	Total obs	150	

Note that the header information indicates that the model is a truncated specification, and that the sample information at the bottom of the screen shows that there are no left and right censored observations.

## Count Models

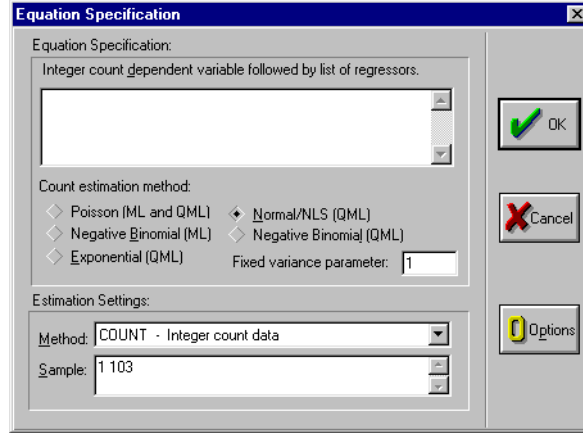
Count models are employed when  $y$  takes integer values that represent the number of events that occur—examples of count data include the number of patents filed by a company, and the number of spells of unemployment experienced over a fixed time interval.

EViews provides support for the estimation of several models of count data. In addition to the standard poisson and negative binomial maximum likelihood (ML) specifications, EViews provides a number of quasi-maximum likelihood (QML) estimators for count data.

### Estimating Count Models in EViews

To estimate a count data model, select **Quick/Estimate Equation...** from the main menu, and select **COUNT** as the estimation method. EViews displays the count estimation dialog into which

you will enter the dependent and explanatory variable regressors, select a type of count model, and if desired, set estimation options:



There are three parts to the specification of the count model:

- In the upper edit field, you should list the dependent variable and the independent variables. You must specify your model by list. The list of explanatory variables specifies a model for the conditional mean of the dependent variable:

$$m(x_i, \beta) = E(y_i | x_i, \beta) = \exp(x_i' \beta). \quad (17.38)$$

- Next, click on **Options** and, if desired, change the default estimation algorithm, convergence criterion, starting values, and method of computing the coefficient covariance.
- Lastly, select one of the entries listed under count estimation method, and if appropriate, specify a value for the variance parameter. Details for each method are provided in the following discussion.

## Poisson Model

For the Poisson model, the conditional density of  $y_i$  given  $x_i$  is

$$f(y_i | x_i, \beta) = \frac{e^{-m(x_i, \beta)} m(x_i, \beta)^{y_i}}{y_i!}. \quad (17.39)$$

where  $y_i$  is a non-negative integer valued random variable. The maximum likelihood estimator (MLE) of the parameter  $\beta$  is obtained by maximizing the log likelihood function

$$l(\beta) = \sum_{i=1}^N y_i \log m(x_i, \beta) - m(x_i, \beta) - \log(y_i!) \quad (17.40)$$

Provided the conditional mean function is correctly specified and the conditional distribution of  $y$  is Poisson, the MLE  $\hat{\beta}$  is consistent, efficient, and asymptotically normally distributed, with variance matrix consistently estimated by

$$\hat{V} = \text{var}(\hat{\beta}) = \left( \sum_{i=1}^N \left( \frac{\partial \hat{m}_i}{\partial \beta} \frac{\partial \hat{m}_i}{\partial \beta} \right) / \hat{m}_i \right), \quad (17.41)$$

where  $\hat{m}_i = m(x_i, \hat{\beta})$ .

The Poisson assumption imposes restrictions that are often violated in empirical applications. The most important restriction is the equality of the (conditional) mean and variance:

$$v(x_i, \beta) = \text{var}(y_i | x_i, \beta) = E(y_i | x_i, \beta) = m(x_i, \beta). \quad (17.42)$$

If the mean-variance equality does not hold, the model is misspecified. EViews provides a number of other estimators for count data which relax this restriction.

We note here that the Poisson estimator may also be interpreted as a quasi-maximum likelihood estimator. The implications of this result are discussed below.

### Negative binomial (ML)

One common alternative to the Poisson model is to estimate the parameters of the model using maximum likelihood of a negative binomial specification. The log likelihood for the negative binomial distribution is given by

$$\begin{aligned} l(\beta, \eta) = & \sum_{i=1}^N y_i \log(\eta^2 m(x_i, \beta)) - (y_i + \eta^2) \log(1 + \eta^2 m(x_i, \beta)) \\ & + \log \Gamma(y_i + 1/\eta^2) - \log(y_i!) - \log \Gamma(1/\eta^2) \end{aligned} \quad (17.43)$$

where  $\eta^2$  is a variance parameter to be jointly estimated with the conditional mean parameters  $\beta$ . Standard errors are computed using the inverse of the information matrix.

The negative binomial distribution is often used when there is *overdispersion* in the data, so that  $v(x_i, \beta) > m(x_i, \beta)$ , since the following moment conditions hold:

$$\begin{aligned} E(y_i | x_i, \beta) &= m(x_i, \beta) \\ \text{var}(y_i | x_i, \beta) &= m(x_i, \beta)(1 + \eta^2 m(x_i, \beta)) \end{aligned} \quad (17.44)$$

$\eta^2$  is therefore a measure of the extent to which the conditional variance exceeds the conditional mean.

Consistency and efficiency of the negative binomial ML requires that the conditional distribution of  $y$  be negative binomial.

## Quasi-maximum likelihood (QML)

We can perform maximum likelihood estimation under a number of alternative distributional assumptions. These *quasi-maximum likelihood (QML) estimators* are robust in the sense that they produce consistent estimates of the parameters of a correctly specified conditional mean, even if the distribution is incorrectly specified.

This robustness result is exactly analogous to the situation in ordinary regression, where the normal ML estimator (least squares) is consistent, even if the underlying error distribution is not normally distributed. In ordinary least squares, all that is required for consistency is a correct specification of the conditional mean  $m(x_i, \beta) = x_i' \beta$ . For QML count models, all that is required for consistency is a correct specification of the conditional mean  $m(x_i, \beta)$ .

The estimated standard errors computed using the inverse of the information matrix will not, however, be consistent unless the conditional distribution of  $y$  is correctly specified. However, it is possible to estimate the standard errors in a robust fashion so that we can conduct valid inference, even if the distribution is incorrectly specified.

EViews provides options to compute two types of robust standard errors. Click **Options** in the Equation Specification dialog box and mark the **Robust Covariance** option. The **Huber/White** option computes QML standard errors, while the **GLM** option computes standard errors corrected for overdispersion. See “Technical Notes” on page 427 for details on these options.

Further details on QML estimation are provided by Gourioux, Monfort, and Trognon (1994a, 1994b). Wooldridge (1996) provides an excellent summary of the use of QML techniques in estimating parameters of count models. See also the extensive related literature on Generalized Linear Models (McCullagh and Nelder, 1989).

## Poisson

The Poisson MLE is also a QMLE for data from alternative distributions. Provided that the conditional mean is correctly specified, it will yield consistent estimates of the parameters  $\beta$  of the mean function. By default, EViews reports the ML standard errors. If you wish to compute the QML standard errors, you should click on **Options**, select **Robust Covariances**, and select the desired covariance matrix estimator.

## Exponential

The log likelihood for the exponential distribution is given by

$$l(\beta) = \sum_{i=1}^N -\log m(x_i, \beta) - y_i / (m(x_i, \beta)). \quad (17.45)$$

As with the other QML estimators, the exponential QMLE is consistent even if the conditional distribution of  $y_i$  is not exponential, provided that  $m_i$  is correctly specified. By default, EViews reports the robust QML standard errors.

### Normal

The log likelihood for the normal distribution is

$$l(\beta) = \sum_{i=1}^N -\frac{1}{2} \left( \frac{y_i - m(x_i, \beta)}{\sigma} \right)^2 - \frac{1}{2} \log(\sigma^2) - \frac{1}{2} \log(2\pi). \quad (17.46)$$

For *fixed*  $\sigma^2$  and correctly specified  $m_i$ , maximizing the normal log likelihood function provides consistent estimates even if the distribution is not normal. Note that maximizing the normal log likelihood for a fixed  $\sigma^2$  is equivalent to minimizing the sum of squares for the nonlinear regression model:

$$y_i = m(x_i, \beta) + \epsilon_i. \quad (17.47)$$

EViews sets  $\sigma^2 = 1$  by default. You may specify any other (positive) value for  $\sigma^2$  by changing the number in the **Fixed variance parameter** field box. By default, EViews reports the robust QML standard errors when estimating this specification.

### Negative binomial

If we maximize the negative binomial log likelihood, given above, for *fixed*  $\eta^2$ , we obtain the QMLE of the conditional mean parameters  $\beta$ . This QML estimator is consistent even if the conditional distribution of  $y$  is not negative binomial, provided that  $m_i$  is correctly specified.

EViews sets  $\eta^2 = 1$  by default, which is a special case known as the geometric distribution. You may specify any other (positive) value by changing the number in the **Fixed variance parameter** field box. For the negative binomial QMLE, EViews by default reports the robust QMLE standard errors.

## Views of Count Models

EViews provides a full complement of views of count models. You can examine the estimation output, compute frequencies for the dependent variable, view the covariance matrix, or perform coefficient tests. Additionally, you can select **View/Actual, Fitted, Residual...** and pick from a number of views describing the ordinary residuals  $e_{oi} = y_i - m(x_i, \hat{\beta})$ , or you can examine the correlogram and histogram of these residuals. For the most part, all of these views are self-explanatory.

Note, however, that the LR test statistics presented in the summary statistics at the bottom of the equation output, or as computed under the **View/Coefficient Tests/Redundant Variables - Likelihood Ratio...**, have a known asymptotic distribution only if the conditional distribution is correctly specified. Under the weaker GLM assumption that the true variance is proportional to the nominal variance, we can form a quasi-likelihood ratio,  $QLR = LR / \hat{\sigma}^2$ , where  $\hat{\sigma}^2$  is the estimated proportional variance factor. This QLR statistic has an asymptotic  $\chi^2$  distribution under the assumption that the mean is correctly specified and that the variances follow the GLM



structure. EViews does not compute the QLR statistic, but it can be estimated by computing an estimate of  $\hat{\sigma}^2$  based upon the standardized residuals. We provide an example of the use of the QLR test statistic below.

If the GLM assumption does not hold, then there is no usable QLR test statistic with a known distribution; see Wooldridge (1996).

## Procedures for Count Models

Most of the procedures are self-explanatory. Some details are required for the forecasting and residual creation procedures.

**Forecast...** provides you the option to forecast the dependent variable  $y_i$  or the predicted linear index  $x_i'\hat{\beta}$ . Note that for all of these models the forecasts of  $y_i$  are given by  $\hat{y}_i = m(x_i, \hat{\beta})$  where  $m(x_i, \hat{\beta}) = \exp(x_i'\hat{\beta})$

**Make Residual Series...** provides the following three types of residuals for count models:

Ordinary	$e_{oi} = y_i - m(x_i, \hat{\beta})$
Standardized (Pearson)	$e_{si} = \frac{y_i - m(x_i, \hat{\beta})}{\sqrt{v(x_i, \hat{\beta}, \hat{\gamma})}}$
Generalized	$e_g = (\text{varies})$

where the  $\gamma$  are any additional parameters in the variance specification. Note that the specification of the variances may vary significantly between specifications. For example, the Poisson model has  $v(x_i, \hat{\beta}) = m(x_i, \hat{\beta})$ , while the exponential has  $v(x_i, \hat{\beta}) = m(x_i, \hat{\beta})^2$ .

The generalized residuals can be used to obtain the score vector by multiplying the generalized residuals by each variable in  $x$ . These scores can be used in a variety of LM or conditional moment tests for specification testing; see Wooldridge (1996).

## Demonstrations

### A specification test for overdispersion

Consider the model

$$\text{NUMB}_i = \beta_1 + \beta_2 \text{IP}_i + \beta_3 \text{FEB}_i + \epsilon_i, \quad (17.48)$$

where the dependent variable NUMB is the number of strikes, IP is a measure of industrial production, and FEB is a February dummy variable, as reported in Kennan (1985, Table 1).

The results from Poisson estimation of this model are presented below:

Dependent Variable: NUMB  
Method: ML/QML - Poisson Count  
Date: 09/14/97 Time: 10:58  
Sample: 1 103  
Included observations: 103  
Convergence achieved after 4 iterations  
Covariance matrix computed using second derivatives

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	1.725630	0.043656	39.52764	0.0000
IP	2.775334	0.819104	3.388254	0.0007
FEB	-0.377407	0.174520	-2.162539	0.0306

R-squared	0.064502	Mean dependent var	5.495146
Adjusted R-squared	0.045792	S.D. dependent var	3.653829
S.E. of regression	3.569190	Akaike info criterion	5.583421
Sum squared resid	1273.912	Schwarz criterion	5.660160
Log likelihood	-284.5462	Hannan-Quinn criter.	5.614503
Restr. log likelihood	-292.9694	Avg. log likelihood	-2.762584
LR statistic (2 df)	16.84645	LR index (Pseudo-R2)	0.028751
Probability(LR stat)	0.000220		

Cameron and Trivedi (1990) propose a regression based test of the Poisson restriction  $v(x_i, \beta) = m(x_i, \beta)$ . To carry out the test, first estimate the Poisson model and obtain the fitted values of the dependent variable. Click **Forecast** and provide a name for the forecasted dependent variable, say NUMB\_F. The test is based on an auxiliary regression of  $e_{oi}^2 - y_i$  on  $\hat{y}_i^2$  and testing the significance of the regression coefficient. For this example, the test regression can be estimated by the command

```
ls (numb-numb_f)^2-numb numb_f^2
```

yielding the following results:

Dependent Variable: (NUMB-NUMB\_F)^2-NUMB  
Method: Least Squares  
Date: 09/14/97 Time: 11:05  
Sample: 1 103  
Included observations: 103

Variable	Coefficient	Std. Error	t-Statistic	Prob.
NUMB_F^2	0.238874	0.052115	4.583571	0.0000

R-squared	0.043930	Mean dependent var	6.872929
Adjusted R-squared	0.043930	S.D. dependent var	17.65726
S.E. of regression	17.26506	Akaike info criterion	8.544908
Sum squared resid	30404.41	Schwarz criterion	8.570488
Log likelihood	-439.0628	Durbin-Watson stat	1.711805

The  $t$ -statistic of the coefficient is highly significant, leading us to reject the Poisson restriction. Moreover, the estimated coefficient is significantly positive, indicating overdispersion in the residuals.

An alternative approach, suggested by Wooldridge (1996) is to regress  $e_{si} - 1$ , on  $\hat{y}_i$ . To perform this test, select **Procs/Make Residual Series...** and select **Standardized**. Save the results in a series, say SRESID. Then estimating the regression specification:

```
sresid^2-1 numbf
```

yields the results:

Dependent Variable: SRESID^2-1				
Method: Least Squares				
Date: 10/06/97 Time: 16:05				
Sample: 1 103				
Included observations: 103				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
NUMBF	0.221238	0.055002	4.022326	0.0001
R-squared	0.017556	Mean dependent var		1.161573
Adjusted R-squared	0.017556	S.D. dependent var		3.138974
S.E. of regression	3.111299	Akaike info criterion		5.117619
Sum squared resid	987.3786	Schwarz criterion		5.143199
Log likelihood	-262.5574	Durbin-Watson stat		1.764537

Both tests suggest the presence of overdispersion, with the variance approximated by  $v \equiv m(1 + 0.23m)$ .

Given the evidence of overdispersion and the rejection of the Poisson restriction, we will re-estimate the model, allowing for mean-variance inequality. Our approach will be to estimate the two-step negative binomial QMLE specification (termed the *quasi-generalized pseudo-maximum likelihood estimator* by Gourieroux, Monfort, and Trognon (1984a, b)) using the estimate of  $\hat{\eta}^2$  derived above. To compute this estimator, simply select **Negative Binomial (QML)** and enter .22124 in the edit field for **Fixed variance parameter**.

We will use GLM variance calculations so you should click on **Option** in the Equation Specification dialog and mark the **Robust Covariance** and **GLM** options. The estimation results are shown below:

Dependent Variable: NUMB  
 Method: QML - Negative Binomial Count  
 Date: 10/11/97 Time: 23:53  
 Sample: 1 103  
 Included observations: 103  
 QML parameter used in estimation: 0.22124  
 Convergence achieved after 3 iterations  
 GLM Robust Standard Errors & Covariance  
 Variance factor estimate = 2.465660162  
 Covariance matrix computed using second derivatives

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	1.724906	0.102543	16.82135	0.0000
IP	2.833103	1.919447	1.475999	0.1399
FEB	-0.369558	0.377376	-0.979285	0.3274
R-squared	0.064374	Mean dependent var	5.495146	
Adjusted R-squared	0.045661	S.D. dependent var	3.653829	
S.E. of regression	3.569435	Akaike info criterion	5.174385	
Sum squared resid	1274.087	Schwarz criterion	5.251125	
Log likelihood	-263.4808	Hannan-Quinn criter.	5.205468	
Restr. log likelihood	-522.9973	Avg. log likelihood	-2.558066	
LR statistic (2 df)	519.0330	LR index (Pseudo-R2)	0.496210	
Probability(LR stat)	0.000000			

The header indicates that the estimated GLM variance factor is 2.4, suggesting that the negative binomial ML would not have been an appropriate specification. Nevertheless, the negative binomial QML should be consistent, and under the GLM assumption, the standard errors should be consistently estimated. It is worth noting that the coefficients on IP and FEB, which were strongly statistically significant in the Poisson specification, are no longer significantly different from zero at conventional significance levels.

### Quasi-likelihood ratio statistic

As described by Wooldridge (1996), specification testing using likelihood ratio statistics requires some care when based upon QML models. We illustrate here the differences between a standard LR test for significant coefficients and the corresponding QLR statistic.

From the results above, we know that the overall likelihood ratio statistic for the Poisson model is 16.85, with a corresponding  $p$ -value of .0002. This statistic is valid under the assumption that  $m(x_i, \beta)$  is specified correctly and that the mean-variance equality holds.

We can decisively reject the latter hypothesis, suggesting that we should derive the QML estimator with consistently estimated covariance matrix under the GLM variance assumption. While EViews does not automatically adjust the LR statistic to reflect the QML assumption, it is easy enough to compute the adjustment by hand. Following Wooldridge, we construct the QLR statistic by dividing the original LR statistic by the estimated GLM variance factor.

Suppose that the estimated QML equation is named EQ1. Then you can use EViews to compute  $p$ -value associated with this statistic, placing the results in scalars using the following commands:

```
scalar qlr = eq1.@logl/2.226420477
scalar qpval = 1-@cchisq(qlr, 2)
```

You can examine the results by clicking on the scalar objects in the workfile window and viewing the results in the status line. The QLR statistic is 7.5666, and the  $p$ -value is .023. The statistic and  $p$ -value are valid under the weaker conditions that the conditional mean is correctly specified, and that the conditional variance is proportional (but not necessarily equal) to the conditional mean.

## Commands

The following are some examples of carrying out the above procedures using commands.

To estimate a logit of LFP on a constant, AGE, EDU, and EDU squared, type

```
equation eq1.binary(d=1) lfp c age edu edu^2
```

To carry out the Hosmer-Lemeshow goodness-of-fit test for the estimated equation EQ1, type

```
eq1.fittest(h)
```

To save the generalized residuals of EQ1 as named series RES\_G, type

```
eq1.makeresid(g) res_g
```

To forecast (static) from EQ1 the index  $x_i'\beta$  as named series XB, type

```
eq1.fit(i) xb
```

For a complete list of commands and options available for estimation methods in this chapter, see the *Command and Programming Reference*.

## Technical Notes

### Huber/White (QML) standard errors

The Huber/White options for robust standard errors computes the quasi-maximum likelihood (or pseudo-ML) standard errors

$$\text{var}_{QML}(\hat{\beta}) = \hat{H}^{-1} \hat{g} \hat{g}' \hat{H}^{-1}, \quad (17.49)$$

where  $\hat{g}$  and  $\hat{H}^{-1}$  are the gradient (or score) and Hessian of the log likelihood evaluated at the ML estimates.

Note that these standard errors are *not* robust to heteroskedasticity in binary dependent variable models. They are robust to certain misspecifications of the underlying distribution of  $y$ .

### GLM standard errors

Many of the discrete and limited dependent variable models described in this chapter belong to a class of models known as *generalized linear models* (GLM). The assumption of GLM is that the distribution of the dependent variable  $y_i$  belongs to the exponential family and that the conditional mean of  $y_i$  is a (smooth) nonlinear transformation of the linear part  $x_i'\beta$ :

$$E(y_i|x_i, \beta) = h(x_i'\beta). \quad (17.50)$$

Even though the QML covariance is robust to general misspecification of the conditional distribution of  $y_i$ , it does not possess any efficiency properties. An alternative consistent estimate of the covariance is obtained if we impose the GLM condition that the (true) variance of  $y_i$  is proportional to the variance of the distribution used to specify the log likelihood:

$$\text{var}(y_i|x_i, \beta) = \sigma^2 \text{var}_{ML}(y_i|x_i, \beta). \quad (17.51)$$

In other words, the ratio of the (conditional) variance to the mean is some constant  $\sigma^2$  that is independent of  $x$ . The most empirically relevant case is  $\sigma^2 > 1$ , which is known as *overdispersion*. If this proportional variance condition holds, a consistent estimate of the GLM covariance is given by

$$\text{var}_{GLM}(\hat{\beta}) = \hat{\sigma}^2 \text{var}_{ML}(\hat{\beta}), \quad (17.52)$$

where

$$\hat{\sigma}^2 = \frac{1}{N-K} \cdot \sum_{i=1}^N \frac{(y_i - \hat{y}_i)^2}{\sqrt{v(x_i, \hat{\beta}, \hat{\gamma})}} = \frac{1}{N-K} \cdot \sum_{i=1}^N \frac{\hat{u}_i^2}{\sqrt{(v(x_i, \hat{\beta}, \hat{\gamma}))}}. \quad (17.53)$$

If you select GLM standard errors, the estimated proportionality term  $\hat{\sigma}^2$  is reported as the variance factor estimate in EViews.

For more discussion on GLM and the phenomenon of overdispersion, see McCullough and Nelder (1989) or Fahrmeir and Tutz (1994).

### The Hosmer-Lemeshow Test

Let the data be grouped into  $j = 1, \dots, J$  groups, and let  $n_j$  be the number of observations in group  $j$ . Define the number of  $y_i = 1$  observations and the average of predicted values in group  $j$  as

$$\begin{aligned} y(j) &= \sum_{i \in j} y_i \\ \bar{p}(j) &= \sum_{i \in j} \hat{p}_i / n_j = \sum_{i \in j} (1 - F(-x_i'\hat{\beta})) / n_j \end{aligned} \quad (17.54)$$

The Hosmer-Lemeshow test statistic is computed as

$$HL = \sum_{j=1}^J \frac{(y(j) - n_j \bar{p}(j))^2}{n_j \bar{p}(j)(1 - \bar{p}(j))}. \quad (17.55)$$

The distribution of the HL statistic is not known, however Hosmer and Lemeshow (1989, p.141) report that extensive simulation indicates that the distribution when the model is correct is well approximated by a  $\chi^2$  distribution with  $J - 2$  degrees of freedom. Note that these findings are based on a simulation where  $J$  is close to  $n$ .

### The Andrews Test

Let the data be grouped into  $j = 1, \dots, J$  groups. Since  $y$  is binary, there are  $2J$  cells into which any observation can fall. Andrews (1988) compares the  $2J$  vector of the actual number of observations in each cell to those predicted from the model, forms a quadratic form, and shows that the quadratic form has an asymptotic  $\chi^2$  distribution if the model is specified correctly.

Andrews suggests three tests depending on the choice of the weighting matrix in the quadratic form. EViews uses the test that can be computed by an auxiliary regression as described in Andrews (1988a, 3.18) or Andrews (1988b, 17).

Briefly, let  $\tilde{A}$  be an  $n \times J$  matrix with element  $\tilde{a}_{ij} = 1(i \in j) - \hat{p}_i$ , where the indicator function  $1(i \in j)$  takes the value one if observation  $i$  belongs to group  $j$  with  $y_i = 1$ , and zero otherwise. (We drop the columns for the groups with  $y = 0$  to avoid singularity.) Let  $B$  be the  $n \times K$  matrix of the contributions to the score  $\partial l(\beta) / \partial \beta'$ . The Andrews test statistic is  $n$  times the  $R^2$  from regressing a constant (one) on each column of  $\tilde{A}$  and  $B$ . Under the null hypothesis that the model is correctly specified,  $nR^2$  is asymptotically distributed  $\chi^2$  with  $J$  degrees of freedom.





## Chapter 18. The Log Likelihood (LogL) Object

---

EViews contains customized procedures which help solve the majority of the estimation problems that you might encounter. On occasion, however, you may come across an estimation specification which is not included among these specialized routines. This specification may be an extension of an existing procedure, or it could be an entirely new class of problem.

Fortunately, EViews provides you with tools to estimate a wide variety of specifications through the *log likelihood (LogL)* object. The LogL object provides you with a general, open-ended tool for estimating a broad class of specifications by maximizing a likelihood function with respect to parameters.

When working with a log likelihood object, you will use EViews' series generation capabilities to describe the log likelihood contribution of each observation in your sample as a function of unknown parameters. You may supply analytical derivatives of the likelihood for one or more parameters, or you can simply let EViews calculate numeric derivatives automatically. EViews will search for the parameter values that maximize the specified likelihood function, and will provide estimated standard errors for these parameter estimates.

In this chapter we provide an overview and describe the general features of the LogL object. We also give examples of specifications which may be estimated using the object. The examples include: multinomial logit, unconditional maximum likelihood AR(1) estimation, Box-Cox regression, disequilibrium switching models, least squares with multiplicative heteroskedasticity, probit specifications with heteroskedasticity, probit with grouped data, nested logit, zero-altered Poisson models, Heckman sample selection models, Weibull hazard models, GARCH(1,1) with t-distributed errors, GARCH with coefficient restrictions, and EGARCH with a generalized error distribution.

### Overview

Most of the work in estimating a model using the LogL object is in creating the text specification which will be used to evaluate the likelihood function.

If you are familiar with the process of generating series in EViews, you should find it relatively easy to work with the LogL specification, since the likelihood specification is merely a list of series assignment statements which are evaluated iteratively during the course of the maximization procedure. All you need to do is write down a set of state-

ments which, when evaluated, will describe a series containing the contributions of each observation to the log likelihood function.

To take a simple example, suppose you believe that your data are generated by the conditional heteroskedasticity regression model:

$$\begin{aligned} y_t &= \beta_1 + \beta_2 x_t + \beta_3 z_t + \epsilon_t \\ \epsilon_t &\sim N(0, \sigma^2 z_t^\alpha) \end{aligned} \quad (18.1)$$

where  $x$ ,  $y$ , and  $z$  are the observed series (data) and  $\beta_1, \beta_2, \beta_3, \sigma, \alpha$  are the parameters of the model. The log likelihood function (the log of the density of the observed data) for a sample of  $T$  observations can be written as

$$\begin{aligned} l(\beta, \alpha, \sigma) &= -\frac{T}{2}(\log(2\pi) + \log \sigma^2) - \frac{\alpha}{2} \sum_{t=1}^T \log(z_t) - \sum_{t=1}^T \frac{(y_t - \beta_1 - \beta_2 x_t - \beta_3 z_t)^2}{z_t^\alpha} \\ &= \sum_{t=1}^T \left\{ \log \phi\left(\frac{y_t - \beta_1 - \beta_2 x_t - \beta_3 z_t}{\sigma z_t^{\alpha/2}}\right) - \frac{1}{2} \log(\sigma^2 z_t^\alpha) \right\} \end{aligned} \quad (18.2)$$

where  $\phi$  is the standard normal density function.

Note that we can write the log likelihood function as a sum of the log likelihood contributions for each observation  $t$ :

$$l(\beta, \alpha, \sigma) = \sum_{t=1}^T l_t(\beta, \alpha, \sigma) \quad (18.3)$$

where the individual contributions are given by

$$l_t(\beta, \alpha, \sigma) = \log \phi\left(\frac{y_t - \beta_1 - \beta_2 x_t - \beta_3 z_t}{\sigma z_t^{\alpha/2}}\right) - \frac{1}{2} \log(\sigma^2 z_t^\alpha) \quad (18.4)$$

Suppose that you know the true parameter values of the model, and you wish to generate a series in EViews which contains the contributions for each observation. To do this, you could assign the known values of the parameters to the elements C(1) to C(5) of the coefficient vector, and then execute the following list of assignment statements as commands or in an EViews program:

```
series res = y - c(1) - c(2)*x - c(3)*z
series var = c(4) * z^c(5)
series logl1 = log(@dnorm(res/@sqrt(var))) - log(var)/2
```

The first two statements describe series which will contain intermediate results used in the calculations. The first statement creates the residual series, RES, and the second statement creates

the variance series, VAR. The series LOGL1 contains the set of log likelihood contributions for each observation.

Now suppose instead that you do not know the true parameter values of the model, and would like to estimate them from the data. The maximum likelihood estimates of the parameters are defined as the set of parameter values which produce the largest value of the likelihood function evaluated across all the observations in the sample.

The LogL object makes finding these maximum likelihood estimates easy. Simply create a new log likelihood object, input the assignment statements above into the LogL specification view, then ask EViews to estimate the specification.

In entering the assignment statements, you need only make two minor changes to the text above. First, the `series` keyword must be removed from the beginning of each line (since the likelihood specification implicitly assumes it is present). Second, an extra line must be added to the specification which identifies the name of the series in which the likelihood contributions will be contained. Thus, you should enter the following into your log likelihood object:

```
@logl logl1
res = y - c(1) - c(2)*x - c(3)*z
var = c(4) * z^c(5)
logl1 = log(@dnorm(res/@sqrt(var))) - log(var)/2
```

The first line in the log likelihood specification, `@logl logl1`, tells EViews that the series LOGL1 should be used to store the likelihood contributions. The remaining lines describe the computation of the intermediate results, and the actual likelihood contributions.

When you tell EViews to estimate the parameters of this model, it will execute the assignment statements in the specification repeatedly for different parameter values, using an iterative algorithm to search for the set of values that maximize the sum of the log likelihood contributions. When EViews can no longer improve the overall likelihood, it will stop iterating and will report final parameter values and estimated standard errors in the estimation output.

The remainder of this chapter discusses the rules for specification, estimation and testing using the likelihood object in greater detail.

## Specification

To create a likelihood object, choose **Objects/New Object.../LogL** or type `logl` in the command window. The likelihood window will open with a blank specification view. The specification view is a text window into which you enter a list of statements which describe your statistical model, and in which you set options which control various aspects of the estimation procedure.

## Specifying the Likelihood

As described in the overview above, the core of the likelihood specification is a set of assignment statements which, when evaluated, generate a series containing the log likelihood contribution of each observation in the sample. There can be as many or as few of these assignment statements as you wish.

Each likelihood specification must contain a control statement which provides the name of the series which is used to contain the likelihood contributions. The format of this statement is

```
@logl series_name
```

where `series_name` is the name of the series which will contain the contributions. This control statement may appear anywhere in the LogL specification.

Whenever the specification is evaluated, whether for estimation or for carrying out a View or Proc, each assignment statement will be evaluated at the current parameter values, and the results stored in a series with the specified name. If the series does not exist, it will be created automatically. If the series already exists, EVIEWS will use the existing series for storage, and will overwrite the data contained in the series.

If you would like to remove one or more of the series used in the specification after evaluation, you can use the `@temp` statement:

```
@temp series_name1 series_name2 ...
```

This statement tells EVIEWS to delete any series in the list after evaluation of the specification is completed. Deleting these series may be useful if your LogL creates a lot of intermediate results, and you do not want the series containing these results to clutter your workfile.

## Parameter Names

In the example above, we used the coefficients C(1) to C(5) as names for our unknown parameters. More generally, any element of a named coefficient vector which appears in the specification will be treated as a parameter to be estimated.

In the conditional heteroskedasticity example, you might choose to use coefficients from three different coefficient vectors: one vector for the mean equation, one for the variance equation, and one for the variance parameters. You would first create three named coefficient vectors by the commands

```
coef(3) beta
coef(1) scale
coef(1) alpha
```

You could then write the likelihood specification as

```

@logl logl1
res = y - beta(1) - beta(2)*x - beta(3)*z
var = scale(1)*z^alpha(1)
logl1 = log(@dnorm(res/@sqrt(var))) - log(var)/2

```

Since all elements of named coefficient vectors in the specification will be treated as parameters, you should make certain that all coefficients really do affect the value of one or more of the likelihood contributions. If a parameter has no effect upon the likelihood, you will experience a singularity error when you attempt to estimate the parameters.

Note that all objects other than coefficient elements will be considered fixed and will not be updated during estimation. For example, suppose that SIGMA is a named scalar in your workfile. Then if you redefine the subexpression for VAR as

```
var = sigma*z^alpha(1)
```

EViews will not estimate SIGMA. The value of SIGMA will remain fixed at its value at the start of estimation.

## Order of Evaluation

The LogL specification contains one or more assignment statements which generate the series containing the likelihood contributions. EViews always evaluates from top to bottom when executing these assignment statements, so expressions which are used in subsequent calculations should always be placed first.

EViews must also iterate through the observations in the sample. Since EViews iterates through both the equations in the specification and the observations in the sample, you will need to specify the order in which the evaluation of observations and equations occurs.

By default, EViews evaluates the specification *by observation* so that *all of the assignment statements* are evaluated for the *first observation*, then for the second observation, and so on across all the observations in the estimation sample. This is the correct order for recursive models where the likelihood of an observation depends on previously observed (lagged) values, as in AR or ARCH models.

You can change the order of evaluation so EViews evaluates the specification *by equation*, so the *first assignment statement* is evaluated for *all the observations*, then the second assignment statement is evaluated for all the observations, and so on for each of the assignment statements in the specification. This is the correct order for models where aggregate statistics from intermediate series are used as input to subsequent calculations.

You can explicitly select which method of evaluation you would like by adding a statement to the likelihood specification. To force evaluation by equation, simply add a line containing the key-

word `@byeqn`. To explicitly state that you require evaluation by observation, the `@byobs` keyword can be used. If no keyword is provided, `@byobs` is assumed.

In the conditional heteroskedasticity example above, it does not matter whether the assignment statements are evaluated by equation (line by line) or by observation, since the results do not depend upon the order of evaluation.

However, if the specification has a recursive structure, or if the specification requires the calculation of aggregate statistics based on intermediate series, you must select the appropriate evaluation order if the calculations are to be carried out correctly.

As an example of the `@byeqn` statement, consider the following specification:

```
@logl robust1
@byeqn
res1 = y-c(1)-c(2)*x
delta = @abs(res1)/6/@median(@abs(res1))
weight = (delta<1)*(1-delta^2)^2
robust1 = -(weight*res1^2)
```

This specification performs robust regression by downweighting outlier residuals at each iteration. The assignment statement for `delta` computes the median of the absolute value of the residuals in each iteration, and this is used as a reference point for forming a weighting function for outliers. The `@byeqn` statement instructs `EViews` to compute all residuals `RES1` at a given iteration before computing the median of those residuals when calculating the `DELTA` series.

## Analytic Derivatives

By default, when maximizing the likelihood and forming estimates of the standard errors, `EViews` computes numeric derivatives of the likelihood function with respect to the parameters. If you would like to specify an analytic expression for one or more of the derivatives, you may use the `@deriv` statement. The `@deriv` statement has the form:

```
@deriv pname1 sname1 pname2 sname2 ...
```

where `pname` is a parameter in the model and `sname` is the name of the corresponding derivative series generated by the specification.

For example, consider the following likelihood object that specifies a multinomial logit model. (See Greene (1997), Chapter 19.7.1 for a discussion of multinomial logit models).

```

' multinomial logit with 3 outcomes
@logl logl1
xb2 = b2(1)+b2(2)*x1+b2(3)*x2
xb3 = b3(1)+b3(2)*x1+b3(3)*x2
denom = 1+exp(xb2)+exp(xb3)
' derivatives wrt the 2nd outcome params
@deriv b2(1) grad21 b2(2) grad22 b2(3) grad23
grad21 = d2-exp(xb2)/denom
grad22 = grad21*x1
grad23 = grad21*x2
' derivatives wrt the 3rd outcome params
@deriv b3(1) grad31 b3(2) grad32 b3(3) grad33
grad31 = d3-exp(xb3)/denom
grad32 = grad31*x1
grad33 = grad31*x2
' specify log likelihood
logl1 = d2*xb2+d3*xb3-log(1+exp(xb2)+exp(xb3))

```

There are three possible outcomes, and the parameters of the three regressors (X1, X2 and the constant) are normalized relative to the first outcome. The analytic derivatives are particularly simple for the multinomial logit model and the two `@deriv` statements in the specification instruct EViews to use the expressions for GRAD21, GRAD22, GRAD23, GRAD31, GRAD32, and GRAD33, instead of computing numeric derivatives.

When working with analytic derivatives, you may wish to check the validity of your expressions for the derivatives by comparing them with numerically computed derivatives. EViews provides you with tools which will perform this comparison at the current values of parameters or at the specified starting values. See the discussion of the **Check Derivatives** view of the likelihood object in the Command Reference at the end of this Chapter.

## Derivative Step Sizes

If analytic derivatives are not specified for any of your parameters, EViews numerically evaluates the derivatives of the likelihood function for those parameters. The step sizes used in computing the derivatives are controlled by two parameters:  $r$  (relative step size) and  $m$  (minimum step size). Let  $\theta^{(i)}$  denote the value of the parameter  $\theta$  at iteration  $i$ . Then the step size at iteration  $i + 1$  is determined by

$$s^{(i+1)} = \max(r\theta^{(i)}, m) \quad (18.5)$$

The two-sided numeric derivative is evaluated as

$$\frac{f(\theta^{(i)} + s^{(i+1)}) - f(\theta^{(i)} - s^{(i+1)})}{2s^{(i+1)}} \quad (18.6)$$

while the one-sided numeric derivative is evaluated as

$$\frac{f(\theta^{(i)} + s^{(i+1)}) - f(\theta^{(i)})}{s^{(i+1)}} \quad (18.7)$$

where  $f$  is the likelihood function. Two-sided derivatives are more accurate, but require roughly twice as many evaluations of the likelihood function and so take about twice as long to evaluate.

The `@derivstep` statement can be used to control the step size and method used to evaluate the derivative at each iteration. The `@derivstep` keyword should be followed by sets of three arguments: the name of the parameter to be set (or the keyword `@all`), the relative step size, and the minimum step size.

The default setting is (approximately):

```
@derivstep(1) @all 1.49e-8 1e-10
```

where “1” in the parentheses indicates that one-sided numeric derivatives should be used and `@all` indicates that the following setting applies to all of the parameters. The first number following `@all` is the relative step size and the second number is the minimum step size. The default relative step size is set to the square root of machine epsilon ( $1.49 \times 10^{-8}$ ) and the minimum step size is set to  $m = 10^{-10}$ .

The step size can be set separately for each parameter in a single or in multiple `@derivstep` statements. The evaluation method option specified in parentheses is a global option; it cannot be specified separately for each parameter.

For example, if you include the line

```
@derivstep(2) c(2) 1e-7 1e-10
```

the relative step size for coefficient C(2) will be increased to  $m = 10^{-7}$  and a two-sided derivative will be used to evaluate the derivative. In a more complex example,

```
@derivstep(2) @all 1.49e-8 1e-10 c(2) 1e-7 1e-10 c(3) 1e-5 1e-8
```

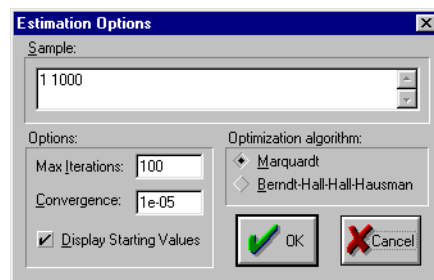
computes two-sided derivatives using the default step sizes for all coefficients except C(2) and C(3). The values for these latter coefficients are specified directly.



## Estimation

Once you have specified the LogL object, you can ask EViews to find the parameter values which maximize the likelihood parameters. Simply click the **Estimate** button in the likelihood window toolbar to open the Estimation Options dialog.

There are a number of options which allow you to control various aspects of the estimation procedure. See Appendix E, “Setting Estimation Options” on page 594 for a discussion of these options. The default settings, however, should provide a good start for most problems. When you click on OK, EViews will begin estimation using the current settings.



### Starting values

Since EViews uses an iterative algorithm to find the maximum likelihood estimates, the choice of starting values is important. For problems in which the likelihood function is globally concave, it will influence how many iterations are taken for estimation to converge. For problems where the likelihood function is not concave, it may determine which of several local maxima is found. In some cases, estimation will fail unless reasonable starting values are provided.

By default, EViews uses the values stored in the coefficient vector or vectors prior to estimation. If a `@param` statement is included in the specification, the values specified in the statement will be used instead.

In our conditional heteroskedasticity regression example, one choice for starting values for the coefficients of the mean equation coefficients are the simple OLS estimates, since OLS provides consistent point estimates even in the presence of (bounded) heteroskedasticity. To use the OLS estimates as starting values, first estimate the OLS equation by the command

```
equation eq1.ls y c x z
```

After estimating this equation, the elements C(1), C(2), C(3) of the C coefficient vector will contain the OLS estimates. To set the variance scale parameter C(4) to the estimated OLS residual variance, you can type the assignment statement in the command window

```
c(4) = eq1.@se^2
```

For the final heteroskedasticity parameter C(5), you can use the residuals from the original OLS regression to carry out a second OLS regression, and set the value of C(5) to the appropriate coefficient. Alternatively, you can arbitrarily set the parameter value using a simple assignment statement:

`c(5) = 1`

Now, if you estimate the LogL specification immediately after carrying out the OLS estimation and subsequent commands, it will use the values that you have placed in the C vector as starting values.

As noted above, an alternative method of initializing the parameters to known values is to include a `@param` statement in the likelihood specification. For example, if you include the line

```
@param c(1) 0.1 c(2) 0.1 c(3) 0.1 c(4) 1 c(5) 1
```

in the specification of the LogL, EViews will always set the starting values to  $C(1) = C(2) = C(3) = 0.1$ ,  $C(4) = C(5) = 1$ .

See also the discussion of starting values in Appendix E, page 595.

## Estimation Sample

EViews uses the sample of observations specified in the Estimation Options dialog when estimating the parameters of the log likelihood. EViews evaluates each expression in the LogL for every observation in the sample at current parameter values, using the *by observation* or *by equation* ordering. All of these evaluations follow the standard EViews rules for evaluating series expressions.

If there are missing values in the log likelihood series at the initial parameter values, EViews will issue an error message and the estimation procedure will stop. In contrast to the behavior of other EViews built-in procedures, LogL estimation performs no endpoint adjustments or dropping of observations with missing values when estimating the parameters of the model.

## LogL Views

- **Likelihood Specification:** displays the window where you specify and edit the likelihood specification.
- **Estimation Output:** displays the estimation results obtained from maximizing the likelihood function.
- **Covariance Matrix:** displays the estimated covariance matrix of the parameter estimates. These are computed from the inverse of the sum of the outer product of the first derivatives evaluated at the optimum parameter values. To save this covariance matrix as a (SYM) MATRIX, you may use the `@cov` function.
- **Wald Coefficient Tests...:** performs the Wald coefficient restriction test. See Chapter 14, Coefficient Tests, for a discussion of Wald tests.

- **Gradient Table:** displays the gradients (first derivatives) of the log likelihood at the current parameter values (if the model has not yet been estimated), or at the converged parameter values (if the model has been estimated). This view may be a useful diagnostic tool if you are experiencing problems with convergence.
- **Check Derivatives:** displays the values of the numeric derivatives and analytic derivatives (if available) at the starting values (if a @param statement is included), or at current parameter values (if there is no @param statement).

## LogL Procs

- **Estimate...:** brings up a dialog to set estimation options, and to estimate the parameters of the log likelihood.
- **Make Model:** creates an untitled model object out of the estimated likelihood specification.
- **Make Gradients:** creates an untitled group of the gradients (first derivatives) of the log likelihood at the estimated parameter values. These gradients are often used in constructing Lagrange multiplier tests.
- **Update Coefs from Loglike:** updates the coefficient vector(s) with the estimates from the likelihood object. This procedure allows you to export the maximum likelihood estimates for use as starting values in other estimation problems.

Most of these Views and Procs should be familiar to you from other EViews estimation objects. We describe below the features that are specific to the LogL object.

## Estimation Output

In addition to the coefficient and standard error estimates, the standard output for the LogL object describes the method of estimation, sample used in estimation, date and time that the LogL was estimated, evaluation order, and information about the convergence of the estimation procedure.

EViews also provides the log likelihood value, average log likelihood value, number of coefficients, and three Information Criteria. By default, the starting values are not displayed. Here we have used the Estimation Options dialog to instruct EViews to display the estimation starting values in the output

LogL: MLOGIT    Workfile: MLOGIT				
View   Procs   Objects   Print   Name   Freeze   MergeText   Estimate   Stats   Spe				
LogL: MLOGIT				
Method: Maximum Likelihood (Marquardt)				
Date: 06/26/98    Time: 14:38				
Sample: 1 1000				
Included observations: 1000				
Evaluation order: By observation				
Initial Values: B2(1)=-1.08356, B2(2)=0.90467, B2(3)=-0.06786, B3(1)=-0.69842, B3(2)=-0.33212, B3(3)=0.32981				
Convergence achieved after 7 iterations				
	Coefficient	Std. Error	z-Statistic	Prob.
B2(1)	-0.521793	0.205568	-2.538302	0.0056
B2(2)	0.994358	0.267963	3.710798	0.0001
B2(3)	0.134983	0.265655	0.508115	0.3057
B3(1)	-0.262307	0.207174	-1.266122	0.1027
B3(2)	0.176770	0.274756	0.643371	0.2600
B3(3)	0.399166	0.274056	1.456511	0.0726
Log likelihood	-1089.415	Akaike info criterion	2.190830	
Avg. log likelihood	-1.089415	Schwarz criterion	2.220277	
Number of Coefs.	6	Hannan-Quinn criter.	2.202022	

## Gradient Table

obs	B2(1)	B2(2)	B2(3)	B3(1)	B3(2)
1	0.617633	0.446035	0.449061	-0.332790	-0.241
2	0.660916	0.310714	0.074241	-0.308290	-0.14
3	0.665234	0.330282	0.518881	-0.355333	-0.171
4	-0.420696	-0.360283	-0.076988	-0.284167	-0.24
5	0.576206	0.516712	0.306708	-0.303744	-0.27
6	-0.394766	-0.296582	-0.147463	0.694417	0.52
7	-0.386750	-0.269690	-0.047604	-0.292850	-0.20
8	0.582613	0.509899	0.362163	-0.311741	-0.27
9	-0.288308	-0.076096	-0.252996	-0.379763	-0.101
10	0.553928	0.549244	0.256273	-0.290475	-0.281
11	-0.301615	-0.086085	-0.056264	0.674758	0.19
12	-0.259970	-0.029439	-0.251003	-0.396446	-0.04
13	0.689356	0.244054	0.363363	-0.346518	-0.12
14					

The **Gradient Table** view displays the gradients (first derivatives) of the log likelihood for every observation. These gradients are computed at the current parameter values (if the model has not yet been estimated), or at the converged parameter values (if the model has been estimated).

You may find this view to be a useful diagnostic tool when experiencing problems with convergence or singularity.

One common problem leading to singular matrices is a zero derivative for a parameter due to an incorrectly specified likelihood, poor starting values, or a lack of model identification. See the discussion below for further details.

## Check Derivatives

You can use the **Check Derivatives** view to examine your numeric derivatives or to check the validity of your expressions for the analytic derivatives. If the LogL specification contains a `@param` statement, the derivatives will be evaluated at the specified values, otherwise, the derivatives will be computed at the current coefficient values.

Coefficient	User	Rel. Step	Min. Step	Coef. Value
B2(1)	GRAD21	1.49E-08	1.00E-10	1.000000
B2(2)	GRAD22	1.49E-08	1.00E-10	1.000000
B2(3)	GRAD23	1.49E-08	1.00E-10	1.000000
B3(1)	GRAD31	1.49E-08	1.00E-10	1.000000
B3(2)	GRAD32	1.49E-08	1.00E-10	1.000000
B3(3)	GRAD33	1.49E-08	1.00E-10	1.000000

Coefficient	Numeric	User	Absolute	Percent
B2(1)	-122.5118	-122.5117	3.65E-08	7.62E-06
B2(2)	-48.24256	-48.24256	-4.32E-08	0.001530
B2(3)	-65.16444	-65.16444	4.28E-08	0.000892
B3(1)	-134.5118	-134.5117	3.65E-08	7.62E-06
B3(2)	-77.25805	-77.25805	-4.13E-08	0.001530
B3(3)	-64.14204	-64.14204	4.28E-08	0.002391

The first part of this view displays the names of the user supplied derivatives, step size parameters, and the coefficient values at which the derivatives are evaluated. The relative and minimum step sizes shown in this example are the default settings.

The second part of the view computes the sum (over all individuals in the sample) of the numeric and, if applicable, the analytic derivatives for each coefficient. If appropriate, EViews will also compute the largest individual difference between the analytic and the numeric derivatives in both absolute, and percentage terms.

## Troubleshooting

Because the LogL object provides a great deal of flexibility, you are more likely to experience problems with estimation using the LogL object than with EViews' built-in estimators.

If you are experiencing difficulties with estimation the following suggestions may help you in solving your problem:

- **Check your likelihood specification.** A simple error involving a wrong sign can easily stop the estimation process from working. You should also verify that the parameters of the model are really identified (in some specifications you may have to impose a normalization across the parameters). Also, every parameter which appears in the model must feed directly or indirectly into the likelihood contributions. The **Check Derivatives** view is particularly useful in helping you spot the latter problem.
- **Choose your starting values.** If any of the likelihood contributions in your sample cannot be evaluated due to missing values or because of domain errors in mathematical operations (logs and square roots of negative numbers, division by zero, etc.) the estimation will stop immediately with the message: "Cannot compute @LogL due to missing values". In other cases, a bad choice of starting values may lead you into regions where the likelihood function is poorly behaved. You should always try to initialize your parameters to sensible numerical values. If you have a simpler estimation technique available which approximates the problem, you may wish to use estimates from this method as starting values for the maximum likelihood specification.
- **Make sure lagged values are initialized correctly.** In contrast to most other estimation routines in EViews, the LogL estimation procedure will not automatically drop observations with NAs or lags from the sample when estimating a log likelihood model. If your likelihood specification involves lags, you will either have to drop observations from the beginning of your estimation sample, or you will have to carefully code the specification so that missing values from before the sample do not cause NAs to propagate through the entire sample (see the AR(1) and GARCH examples for a demonstration).

Since the series used to evaluate the likelihood are contained in your workfile (unless you use the `@temp` statement to delete them), you can examine the values in the log likelihood and intermediate series, to find problems involving lags and missing values.

- **Verify your derivatives.** If you are using analytic derivatives, use the **Check Derivatives** view to make sure you have coded the derivatives correctly. If you are using numerical derivatives, consider specifying analytic derivatives or adjusting the options for derivative method or step size.
- **Reparametrize your model.** If you are having problems with parameter values causing mathematical errors, you may wish to consider reparameterizing the model to restrict the parameter within its valid domain. See the discussion below for examples.

Most of the error messages you are likely to see during estimation are self-explanatory. The error message “near singular matrix” may be less obvious. This error message occurs when EViews is unable to invert the matrix of the sum of the outer product of the derivatives which makes it impossible to determine the direction of the next step of the optimization. This error may indicate a wide variety of problems, including bad starting values, but will almost always occur if the model is not identified, either theoretically, or in terms of the available data.

## Limitations

The likelihood object can be used to estimate parameters that maximize (or minimize) a variety of objective functions. Although the main use of the likelihood object will be to specify a log likelihood, you can specify least squares and minimum distance estimation problems with the likelihood object as long as the objective function is additive over the sample.

You should be aware that the algorithm used in estimating the parameters of the log likelihood is not well suited to solving arbitrary maximization or minimization problems. The algorithm forms an approximation to the Hessian of the log likelihood, based on the sum of the outer product of the derivatives of the likelihood contributions. This approximation relies on both the functional form and statistical properties of maximum likelihood objective functions, and may not be a good approximation in general settings. Consequently, you may or may not be able to obtain results with other functional forms. Furthermore, the standard error estimates of the parameter values will only have meaning if the series describing the log likelihood contributions are (up to an additive constant) the individual contributions to a correctly specified, well-defined theoretical log likelihood.

Currently, the expressions used to describe the likelihood contribution must follow the rules of EViews series expressions. This restriction implies that we do not allow matrix operations in the likelihood specification. In order to specify likelihood functions for multiple equation models, you may have to write out the expression for the determinants and quadratic forms. Although possible, this may become tedious for models with more than two equations.

Additionally, the LogL object does not directly handle optimization subject to general inequality constraints. There are, however, a variety of well-established techniques for imposing simple inequality constraints. We provide examples below. The underlying idea is to apply a monotonic transformation to the coefficient so that the new coefficient term takes on values only in the desired range. The commonly used transformations are the `@exp` for one-sided restrictions and the `@logit` and `@arctan` for two-sided restrictions.

You should be aware of the limitations of the transformation approach. First, the approach only works for relatively simple inequality constraints. If you have several cross-coefficient inequality restrictions, the solution will quickly become intractable. Second, in order to perform hypothesis tests on the untransformed coefficient, you will have to obtain an estimate of the standard errors of the associated expressions. Since the transformations are generally nonlinear, you will have to

compute linear approximations to the variances yourself (using the delta method). Lastly, inference will be poor near the boundary values of the inequality restrictions.

### Simple one-sided restrictions

Suppose you would like to restrict the estimate of the coefficient of  $X$  to be no larger than 1. One way you could do this is to specify the corresponding subexpression as follows:

```
' restrict coef on x to not exceed 1
res1 = y - c(1) - (1-exp(c(2)))*x
```

Note that EViews will report the point estimate and the standard error for the parameter  $C(2)$ , not the coefficient of  $X$ . To find the standard error of the expression  $1 - \exp(c(2))$ , you will have to use the delta method; see for example Greene (1997), Theorems 4.15 and 4.16.

### Simple two-sided restrictions

Suppose instead that you want to restrict the coefficient for  $X$  to be between  $-1$  and  $1$ . Then you can specify the expression as:

```
' restrict coef on x to be between -1 and 1
res1 = y - c(1) - (2*@logit(c(2))-1)*x
```

Again EViews will report the point estimate and standard error for the parameter  $C(2)$ . You will have to use the delta method to compute the standard error of the transformation expression  $2 * \text{@logit}(c(2)) - 1$ .

More generally, if you want to restrict the parameter to lie between  $L$  and  $H$ , you can use the transformation

$$(H-L) * \text{@logit}(c(1)) + L$$

where  $C(1)$  is the parameter to be estimated. In the above example,  $L = -1$  and  $H = 1$ .

## Examples

In this section, we provide extended examples of working with the LogL object to estimate a multinomial logit and a maximum likelihood AR(1) specification. Example programs for these and several other specifications are provided in your default EViews data directory. If you set your default directory to point to the EViews data directory, you should be able to issue a RUN command for each of these programs to create the LogL object and to estimate the unknown parameters.

### Multinomial logit (mlogit1.prg)

In this example, we demonstrate how to specify and estimate a simple multinomial logit model using the LogL object. Suppose the dependent variable  $Y$  can take one of three categories 1, 2,

and 3. Further suppose that there are data on two regressors, X1 and X2 that vary across observations (individuals). Standard examples include variables such as age and level of education. Then the multinomial logit model assumes that the probability of observing each category in Y is given by:

$$\Pr(y_i = j) = \frac{\exp(\beta_{0j} + \beta_{1j}x_{1i} + \beta_{2j}x_{2i})}{\sum_{k=1}^3 \exp(\beta_{0k} + \beta_{1k}x_{1i} + \beta_{2k}x_{2i})} = P_{ij} \quad (18.8)$$

for  $j = 1, 2, 3$ . Note that the parameters  $\beta$  are specific to each category so there are  $3 \times 3 = 9$  parameters in this specification. The parameters are not all identified unless we impose a normalization (see for example Greene, 1997, chapter 19.7), so we normalize the parameters of the first choice category  $j = 1$  to be all zero:  $\beta_{0,1} = \beta_{1,1} = \beta_{2,1} = 0$ .

The log likelihood function for the multinomial logit can be written as

$$l = \sum_{i=1}^N \sum_{j=1}^3 d_{ij} \log(P_{ij}) \quad (18.9)$$

where  $d_{ij}$  is a dummy variable that takes the value 1 if observation  $i$  has chosen alternative  $j$ ; 0 otherwise. The first-order conditions are

$$\frac{\partial l}{\partial \beta_{kj}} = \sum_{i=1}^N (d_{ij} - P_{ij})x_{ki} \quad (18.10)$$

for  $k = 0, 1, 2$  and  $j = 1, 2, 3$ .

We have provided, in the Example Files subdirectory of your default EViews directory, a workfile MLOGIT.WK1 containing artificial multinomial data. The program begins by loading this workfile.

```
' load artificial data
load "c:\eviews3\example files\mlogit"
```

Next, we declare the coefficient vectors that will contain the estimated parameters for each choice alternative.

```
' declare parameter vector
coef(3) b2
coef(3) b3
```

As an alternative, we could have used the default coefficient vector C.

We then set up the likelihood function by issuing a series of append statements:



```

mlogit.append xb2 = b2(1)+b2(2)*x1+b2(3)*x2
mlogit.append xb3 = b3(1)+b3(2)*x1+b3(3)*x2
' define prob for each choice
mlogit.append denom = 1+exp(xb2)+exp(xb3)
mlogit.append pr1 = 1/denom
mlogit.append pr2 = exp(xb2)/denom
mlogit.append pr3 = exp(xb3)/denom
' specify likelihood
mlogit.append logl1 = (1-dd2-dd3)*log(pr1)
                +dd2*log(pr2)+dd3*log(pr3)

```

Since the analytic derivatives for the multinomial logit are particularly simple, we also specify the expressions for the analytic derivatives to be used during estimation and the appropriate `@deriv` statements:

```

' specify analytic derivatives
for !i = 2 to 3
mlogit.append @deriv b{!i}(1) grad{!i}1 b{!i}(2) grad{!i}2
                b{!i}(3) grad{!i}3
mlogit.append grad{!i}1 = dd{!i}-pr{!i}
mlogit.append grad{!i}2 = grad{!i}1*x1
mlogit.append grad{!i}3 = grad{!i}1*x2
next

```

Note that if you were to specify this likelihood interactively, you would simply type the expression that follows each `append` statement directly into the MLOGIT object.

This concludes the actual specification of the likelihood object. Before estimating the model, we get the starting values by estimating a series of binary logit models.

```

' get starting values from binomial logit
equation eq2.binary(d=1) dd2 c x1 x2
b2 = eq2.@coefs
equation eq3.binary(d=1) dd3 c x1 x2
b3 = eq3.@coefs

```

To check, whether you have specified the analytic derivatives correctly, choose **View/Check Derivatives** or use the command

```
show mlogit.checkderiv
```

If you have correctly specified the analytic derivatives, they should be fairly close to the numeric derivatives.

We are now ready to estimate the model. Either click the **Estimate** button or use the command

```
' do MLE
mlogit.ml(d)
show mlogit.output
```

Note that you can examine the derivatives for this model using the **Gradient Table** view, or you can examine the series in the workfile containing the gradients. You can also look at the intermediate results, and log likelihood values. For example, to look at the likelihood contributions for each individual, simply double click on the LOGL1 series.

### AR(1) model (ar1.prg)

In this example, we demonstrate how to obtain full maximum likelihood estimates of an AR(1). The maximum likelihood procedure uses the first observation in the sample, in contrast to the built-in AR(1) procedure in EViews which treats the first observation as fixed and maximizes the conditional likelihood for the remaining observations by nonlinear least squares.

As an illustration, we first generate data that follows an AR(1) process:

```
' make up data
create m 80 89
rndseed 123
series y=0
smpl @first+1 @last
y = 1+0.85*y(-1) + nrnd
```

The exact Gaussian likelihood function for an AR(1) model is given by

$$f(y, \theta) = \begin{cases} \frac{1}{\sigma\sqrt{2\pi(1-\rho^2)}} \exp\left\{-\frac{(y_t - c/(1-\rho^2))^2}{2(\sigma^2/(1-\rho^2))}\right\} & t = 1 \\ \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(y_t - c - \rho y_{t-1})^2}{2(\sigma^2)}\right\} & t > 0 \end{cases} \quad (18.11)$$

where  $c$  is the constant term,  $\rho$  is the AR(1) coefficient, and  $\sigma^2$  is the error variance, all to be estimated (see for example Hamilton, 1994, chapter 5.2).

Since the likelihood function evaluation differs for the first observation in our sample, we create a dummy variable indicator for the first observation:

```

' create dummy variable for first obs
series d1 = 0
smpl @first @first
d1 = 1
smpl @all

```

Next, we declare the coefficient vectors to store the parameter estimates and initialize them with the least squares estimates.

```

' set starting values to LS (drops first obs)
equation eq1.ls y c ar(1)
coef(1) rho = c(2)
coef(1) s2 = eq1.@se^2

```

We then specify the likelihood function. We make use of the `@recode` function to differentiate the evaluation of the likelihood for the first observation from the remaining observations. Note: the `@recode` function used here uses the updated syntax for this function—please double-check the current documentation for details.

```

' set up likelihood
' uses new @recode syntax 6/98
logl ar1
ar1.append @logl logl1
ar1.append var = @recode(d1=1,s2(1)/(1-rho(1)^2),s2(1))
ar1.append res = @recode(d1=1,y-c(1)/(1-rho(1)),y-c(1)-
    rho(1)*y(-1))
ar1.append sres = res/@sqrt(var)
ar1.append logl1 = log(@dnorm(sres))-log(var)/2

```

The likelihood specification uses the built-in function `@dnorm` for the standard normal density. The second term is the Jacobian term that arises from transforming the standard normal variable to one with non-unit variance. (You could of course write out the likelihood for the normal distribution without using the `@dnorm` function.)

The program displays the MLE together with the least squares estimates

```
' do MLE
arl.ml(d)
show arl.output
' compare with EViews AR(1) which ignores first obs
show eql.output
```

## Additional Examples

The following additional example programs can be found in the “Example Files” subdirectory of your default EViews directory.

- **Conditional logit** (clogit1.prg): estimates a conditional logit with 3 outcomes and both individual specific and choice specific regressors. The program also displays the prediction table and carries out a Hausman test for independence of irrelevant alternatives (IIA). See Greene (1997, chapter 19.7) for a discussion of multinomial logit models.
- **Box-Cox transformations** (boxcox1.prg): estimates a simple bivariate regression with an estimated Box-Cox transformation on both the dependent and independent variables. Box-Cox transformation models are notoriously difficult to estimate and the results are very sensitive to starting values.
- **Disequilibrium switching model** (diseq1.prg): estimates the switching model in exercise 15.14–15.15 of Judge et al. (1985, pages 644–646). Note that there are some typos in Judge et al. (1985, pages 639–640). The program uses the likelihood specification in Quandt (1988, page 32, equations 2.3.16–2.3.17).
- **Multiplicative heteroskedasticity** (hetero1.prg): estimates a linear regression model with multiplicative heteroskedasticity. Replicates the results in Greene (1997, example 12.14).
- **Probit with heteroskedasticity** (hprobit1.prg): estimates a probit specification with multiplicative heteroskedasticity. See Greene (1997, example 19.7).
- **Probit with grouped data** (gprobit1.prg): estimates a probit with grouped data (proportions data). Estimates the model in Greene (1997, exercise 19.6).
- **Nested logit** (nlogit1.prg): estimates a nested logit model with 2 branches. Tests the IIA assumption by a Wald test. See Greene (1997, chapter 19.7.4) for a discussion of nested logit models.
- **Zero-altered Poisson model** (zpoiss1.prg): estimates the zero-altered Poisson model. Also carries out the non-nested LR test of Vuong (1989). See Greene (1997, chapter 19.9.6) for a discussion of zero-altered Poisson models and Vuong’s non-nested likelihood ratio test.
- **Heckman sample selection model** (heckman1.prg): estimates Heckman’s two equation sample selection model by MLE using the two-step estimates as starting values.

- **Weibull hazard model** (`weibull1.prg`): estimates the uncensored Weibull hazard model described in Greene (1997, example 20.18). The program also carries out one of the conditional moment tests in Greene (1997, example 20.19).
- **GARCH(1,1) with t-distributed errors** (`arch_t1.prg`): estimates a GARCH(1,1) model with  $t$ -distribution. The log likelihood function for this model can be found in Hamilton (1994, equation 21.1.24, page 662).
- **GARCH with coefficient restrictions** (`garch1.prg`): estimates an MA(1)-GARCH(1,1) model with coefficient restrictions in the conditional variance equation. This model is estimated by Bollerslev, Engle, and Nelson (1994, equation 9.1, page 3015) for different data
- **EGARCH with generalized error distributed errors** (`egarch1.prg`): estimates Nelson's (1991) exponential GARCH with generalized error distribution. The specification and likelihood is described in Hamilton (1994, pages 668–669.)



## Part V. Multiple Equation Analysis

---

Chapters 18–22 document EViews tools for multiple equation estimation, forecasting and data analysis.

- Chapters 18–20 describe estimation techniques for systems of equations, VAR and VEC, and state space models.
- Chapter 21 outlines tools for working with pooled time series, cross-section data, and estimating standard equation specifications which account for the pooled structure of the data.
- Chapter 22 describes the use of model objects to forecast from multiple equation estimates, or to perform multivariate simulation.





## Chapter 19. System Estimation

---

This chapter describes methods of estimating the parameters of systems of equations. We describe least squares, weighted least squares, seemingly unrelated regression (SUR), weighted two-stage least squares, three-stage least squares, full-information maximum likelihood (FIML), and generalized method of moments (GMM) estimation techniques.

Once you have estimated the parameters of your system of equations, you may wish to forecast future values or perform simulations for different values of the explanatory variables. Chapter 23 describes the use of models to forecast from an estimated system of equations or to perform single and multivariate simulation.

### Background

A *system* is a group of equations containing unknown parameters. Systems can be estimated using a number of multivariate techniques that take into account the interdependencies among the equations in the system.

The general form of a system is

$$f(y_t, x_t, \beta) = \epsilon_t, \quad (19.1)$$

where  $y_t$  is a vector of endogenous variables,  $x_t$  is a vector of exogenous variables, and  $\epsilon_t$  is a vector of possibly serially correlated disturbances. The task of estimation is to find estimates of the vector of parameters  $\beta$ .

EViews provides you with a number of methods of estimating the parameters of the system. One approach is to estimate each equation in the system separately, using one of the single equation methods described earlier in this manual. A second approach is to estimate, simultaneously, the complete set of parameters of the equations in the system. The simultaneous approach allows you to place constraints on coefficients across equations and to employ techniques that account for correlation in the residuals across equations.

While there are important advantages to using a system to estimate your parameters, they do not come without cost. Most importantly, if you misspecify one of the equations in the system and estimate your parameters using single equation methods, only the misspecified equation will be poorly estimated. If you employ system estimation techniques, the poor estimates for the misspecification equation may “contaminate” estimates for other equations.

At this point, we take care to distinguish between systems of equations and models. A *model* is a group of known equations describing endogenous variables. Models are used to solve for values of the endogenous variables, given information on other variables in the model.

Systems and models often work together quite closely. You might estimate the parameters of a system of equations, and then create a model in order to forecast or simulate values of the endogenous variables in the system. We discuss this process in greater detail in Chapter 23.

## System Estimation Methods

EViews will estimate the parameters of a system of equations using:

- Ordinary least squares.
- Equation weighted regression.
- Seemingly unrelated regression (SUR).
- Weighted two-stage least squares.
- Three-stage least squares.
- Full information maximum likelihood (FIML).
- Generalized method of moments (GMM).

The equations in the system may be linear or nonlinear, and may contain autoregressive error terms.

In the remainder of this section, we describe each technique at a general level. Users who are interested in the technical details are referred to the “Technical Discussion” on page 468.

### Ordinary Least Squares

This technique minimizes the sum-of-squared residuals for each equation, accounting for any cross-equation restrictions on the parameters of the system. If there are no such restrictions, this method is identical to estimating each equation using single-equation ordinary least squares.

### Cross-Equation Weighting

This method accounts for cross-equation heteroskedasticity by minimizing the weighted sum-of-squared residuals. The equation weights are the inverses of the estimated equation variances, and are derived from unweighted estimation of the parameters of the system. This method yields identical results to unweighted single-equation least squares if there are no cross-equation restrictions.

## Seemingly Unrelated Regression

The seemingly unrelated regression (SUR) method, also known as the multivariate regression, or Zellner's method, estimates the parameters of the system, accounting for heteroskedasticity, and contemporaneous correlation in the errors across equations. The estimates of the cross-equation covariance matrix are based upon parameter estimates of the unweighted system.

Note that EViews estimates a more general form of SUR than is typically described in the literature, since it allows for cross-equation restrictions on parameters.

## Weighted Two-Stage Least Squares

The weighted two-stage least squares (WTLS) estimator is the two-stage version of the weighted least squares estimator described above. WTLS is an appropriate technique when some of the right-hand side variables are correlated with the error terms, and there is heteroskedasticity, but no contemporaneous correlation in the residuals.

EViews first applies TSLS to the unweighted system, enforcing any cross-equation parameter restrictions. The results from this estimation are used to form equation weights, based upon the estimated equation variances. If there are no cross-equation restrictions, these first-stage results will be identical to unweighted single-equation TSLS.

## Three-Stage Least Squares

Three-stage least squares (3SLS) is the two-stage least squares version of the SUR method. It is an appropriate technique when right-hand side variables are correlated with the error terms, and there is both heteroskedasticity, and contemporaneous correlation in the residuals.

EViews applies TSLS to the unweighted system, enforcing any cross-equation parameter restrictions. These estimates are used to form an estimate of the full cross-equation covariance matrix which, in turn, is used to transform the equations to eliminate the cross-equation correlation. TSLS is applied to the transformed model.

## Full Information Maximum Likelihood (FIML)

Full Information Maximum Likelihood (FIML) estimates the likelihood function under the assumption that the contemporaneous errors have a joint normal distribution. Provided that the likelihood function is correctly specified, FIML is fully efficient.

## Generalized Method of Moments (GMM)

The GMM estimator belongs to a class of estimators known as M-estimators that are defined by minimizing some criterion function. GMM is a robust estimator in that it does not require information of the exact distribution of the disturbances.

GMM estimation is based upon the assumption that the disturbances in the equations are uncorrelated with a set of instrumental variables. The GMM estimator selects parameter estimates so

that the correlations between the instruments and disturbances are as close to zero as possible, as defined by a criterion function. By choosing the weighting matrix in the criterion function appropriately, GMM can be made robust to heteroskedasticity and/or autocorrelation of unknown form.

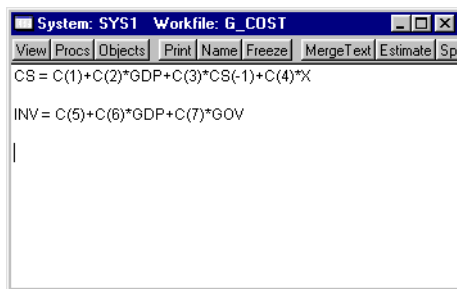
Many standard estimators, including all of the system estimators provided in EViews, can be set up as special cases of GMM. For example, the ordinary least squares estimator can be viewed as a GMM estimator, based upon the conditions that each of the right-hand variables is uncorrelated with the residual.

## How to Create and Specify a System

To estimate the parameters of your system of equations, you should first create a system object and specify the system of equations. Click on **Objects/New Object/System** or type `system` in the command window. The system object window should appear. When you first create the system, the window will be blank. You will fill the system specification window with text describing the equations, and potentially, lines describing the instruments and the parameter starting values.

### Equations

Enter your equations, by formula, using standard EViews expressions. The equations in your system should be behavioral equations with unknown coefficients and an implicit error term.



Consider the specification of a simple two equation system. You can use the default EViews coefficients,  $C(1)$ ,  $C(2)$ , and so on, or you can use other coefficient vectors, in which case you should first declare them by clicking **Objects/New Object/Matrix-Vector-Coef/Coefficient Vector** in the main menu.

There are some general rules for specifying your equations:

- Equations can be nonlinear in their variables, coefficients, or both. Cross equation coefficient restrictions may be imposed by using the same coefficients in different equations. For example,

$$\begin{aligned} y &= c(1) + c(2) * x \\ z &= c(3) + c(2) * z + (1 - c(2)) * x \end{aligned}$$

- You may also impose adding up constraints. Suppose for the equation

$$y = c(1) * x1 + c(2) * x2 + c(3) * x3$$

you wish to impose  $C(1) + C(2) + C(3) = 1$ . You can impose this restriction by specifying the equation as

$$y = c(1) * x1 + c(2) * x2 + (1 - c(1) - c(2)) * x3$$

- The equations in a system may contain autoregressive (AR) error specifications, but not MA, SAR, or SMA error specifications. You must associate coefficients with each AR specification. Enclose the entire AR specification in square brackets and follow each AR with an = sign and a coefficient. For example,

$$cs = c(1) + c(2) * gdp + [ar(1) = c(3), ar(2) = c(4)]$$

You can constrain all of the equations in a system to have the same AR coefficient by giving all equations the same AR coefficient number, or you can estimate separate AR processes, by assigning each equation its own coefficient.

- Equations in a system need not have a dependent variable followed by an equals sign and then an expression. The = sign can be anywhere in the formula, as in:

$$\log(unemp / (1 - unemp)) = c(1) + c(2) * dmr$$

You can also write the equation as a simple expression without a dependent variable, as in

$$(c(1) * x + c(2) * y + 4)^2$$

When encountering an expression that does not contain an equals sign, EViews sets the entire expression equal to the implicit error term.

- If an equation should not have a disturbance, it is an identity, and should not be included in a system. If necessary, you should solve out for any identities to obtain the behavioral equations.
- You should make certain that there is no identity linking all of the disturbances in your system. For example, if each of your equations describes a fraction of a total, the sum of the equations will always equal one, and the sum of the disturbances will identically equal zero. You will need to drop one of these equations to avoid numerical problems.

## Instruments

If you plan to estimate your system using two-stage least squares, three-stage least squares, or GMM, you must specify the instrumental variables to be used in estimation. There are two ways to specify instruments. If you want to use identical instruments in all of your equations, you should include a line beginning with `inst`, followed by a list of all the exogenous variables to be used as instruments. For example the line,

```
inst gdp(-1 to -4) x gov
```

instructs EViews to use these six variables as instruments for all of the equations in the system. If you estimate this model using a method that does not use instruments, the line will be ignored.

You can also specify different instruments for each equation by appending an @ sign at the end of the equation, followed by a list of instruments for that equation. For example,

```
cs=c(1)+c(2)*gdp+c(3)*cs(-1) @ cs(-1) inv(-1) gov
inv=c(4)+c(5)*gdp+c(6)*gov @ gdp(-1) gov
```

The first equation uses CS(-1), INV(-1), GOV, and a constant as instruments, while the second equation uses GDP(-1), GOV, and a constant as instruments.

Lastly, you can mix the two methods. Any equation without individually specified instruments will use the instruments specified by the `inst` statement.

### Additional Comments

- A constant is always included as an instrument in every equation, whether it is specified explicitly or not.
- You should list all exogenous right-hand side variables as instruments for a given equation.
- Identification requires that there should be at least as many instruments (including the constant) in each equation as there are right-hand side variables in that equation.

### Starting Values

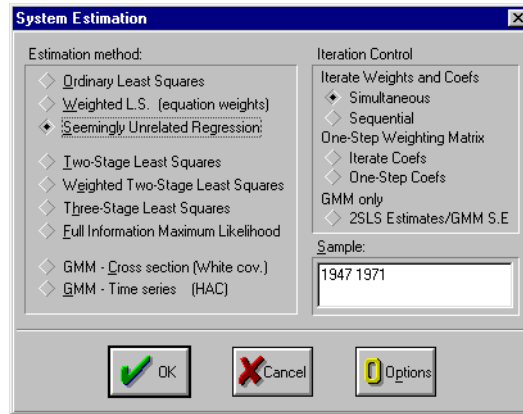
For systems that contain nonlinear equations, you can include a line that begins with `param` to provide starting values for some or all of the parameters. List pairs of parameters and values. For example,

```
param c(1) .15 b(3) .5
```

sets the initial values of C(1) and B(3). If you do not provide starting values, EViews uses the values in the current coefficient vector.

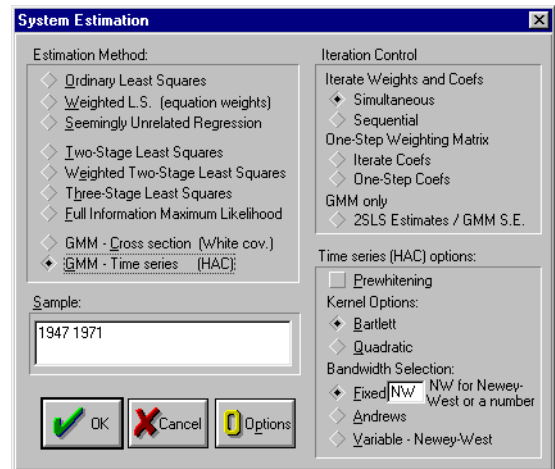
### How to Estimate a System

Once you have created and specified your system, push the **Estimate** button on the toolbar.



First select the estimation method. The **GMM-Cross section** option uses a weighting matrix that is robust to heteroskedasticity and contemporaneous correlation of unknown form, while the **GMM-Time series (HAC)** option extends this robustness to autocorrelation of unknown form.

When you select the **GMM-Time series(HAC)** option, the dialog box expands to display additional options for specifying the weighting matrix. The new options will appear at the lower right part of the dialog control. These options control the computation of the heteroskedasticity and autocorrelation robust (HAC) weighting matrix. See “Technical Discussion” on page 468 for a more detailed discussion of these options.



The **Kernel Options** determines the functional form of the kernel used to weight the autocovariances to compute the weighting matrix. The **Bandwidth Selection** option determines how the weights given by the kernel change with the lags of the autocovariances in the computation of the weighting matrix. If you select **Fixed** bandwidth, you may enter a number for the bandwidth or type **nw** to use Newey and West’s fixed bandwidth selection criterion.

The **Prewhitening** option runs a preliminary VAR(1) prior to estimation to “soak up” the correlation in the moment conditions.

## Iteration Control

For weighted least squares, SUR, weighted TSLS, 3SLS, GMM, and nonlinear systems of equations, there is an additional estimation issue involving the procedure for computing the GLS weighting matrix and the coefficient vector. In general, you can elect to use the EViews default option with no additional difficulties, but there may be times when you wish to exercise greater control over the computation.

The estimation option controls the method of iterating over coefficients, over the weighting matrices, or both:

- **One-Step Weighting Matrix—Iterate Coefs** is the default method.

By default, EViews carries out a first-stage estimation of the coefficients using no weighting matrix (the identity matrix). Using starting values obtained from OLS (or TSLS, if there are instruments), EViews iterates until the coefficients converge. If the model is linear, this procedure involves a single OLS or TSLS regression.

The residuals from this first-stage iteration are used to form a consistent estimate of the weighting matrix.

In the second-stage of the procedure, EViews uses the estimated weighting matrix in forming new estimates of the coefficients. If the model is nonlinear, EViews iterates the coefficient estimates until convergence.

- **One-Step Weighting Matrix—One-Step Coefs** performs the first stage estimation of the coefficients, and constructs an estimate of the weighting matrix. In the second stage, EViews does not iterate the coefficients to convergence, instead performing a single coefficient iteration step.
- **Iterate Weights and Coefs—Sequential** repeats the default method, described above, until both the coefficients and the weighting matrix converge.
- **Iterate Weights and Coefs—Simultaneous** updates both the coefficients and the weighting matrix at each iteration. These steps are then repeated until both the coefficients and weighting matrix converge. This is the iteration method employed in previous versions of EViews.

Note that all four of the estimation techniques yield results that are asymptotically efficient. For linear models, the two **Iterate Weights and Coefs** options are equivalent to each other, and the two **One-Step Weighting Matrix** options are equivalent to each other, since obtaining coefficient estimates does not require iteration.

There is a last technique which is not fully efficient, but which does produce estimated covariances and standard errors which are robust to heteroskedasticity or heteroskedasticity and serial correlation.



- The **2SLS Estimates/GMM S.E.** computes first stage estimates of the parameters assuming an identity weighting matrix. These estimates are then used to compute a coefficient covariance matrix that is robust to cross-section heteroskedasticity (White) or heteroskedasticity and autocorrelation (Newey-West).

## Other Options

The **Options** button in the System Estimation dialog allows you to select an estimation algorithm, change the convergence criterion, or the maximum number of iterations. See Appendix E for additional discussion of these options.

## Estimation Output

The system estimation output contains parameter estimates, standard errors, and  $t$ -statistics for each of the coefficients in the system. Additionally, EViews reports the determinant of the residual covariance matrix, and, for FIML estimates, the maximized likelihood value.

In addition, EViews reports a set of summary statistics for each equation. The  $R^2$  statistic, Durbin-Watson statistic, standard error of the regression, sum-of-squared residuals, etc., are computed for each equation using the standard definitions, based on the residuals from the system estimation procedure.

You may access most of these results using regression statistics functions. See Chapter 11, page 241 for a discussion of the use of these functions, and Appendix A, page 577 for a full listing of the available functions.

## Working With Systems

After obtaining estimates, the system object provides a number of tools for examining the equation results, and performing inference and specification testing.

### System Views

Some of the system views are familiar from the discussion in previous chapters:

- You can examine the estimated covariance matrix by selecting the **Coefficient Covariance Matrix** view.
- **Wald Coefficient Tests...** performs hypothesis tests on the coefficients. These views are discussed in greater depth in Chapter 15, “Specification and Diagnostic Tests”, beginning on page 329.
- The **Estimation Output** view displays the coefficient estimates and summary statistics for the system. You may also access this view by pressing **Stats** on the system toolbar.

Other views are very familiar, but differ slightly in name or output, from their single equation counterparts:

- **System Specification** displays the specification window for the system. The specification window may also be displayed by pressing **Spec** on the toolbar.
- **Residual Graphs** displays a separate graph of the residuals from each equation in the system.
- **Endogenous Table** presents a spreadsheet view of the endogenous variables in the system.
- **Endogenous Graph** displays graphs of each of the endogenous variables.

The last two views are specific to systems:

- **Residual Correlation Matrix** computes the contemporaneous correlation matrix for the residuals of each equation.
- **Residual Covariance Matrix** computes the contemporaneous covariance matrix for the residuals. See also `@residcova` in Appendix A.

## System Procs

One notable difference between systems and single equation objects is that there is no forecast procedure for systems. To forecast or perform simulation using an estimated system, you must use a model object.

EViews provides you with a simple method of incorporating the results of a system into a model. If you select **Procs/Make Model**, EViews will open an untitled model object containing the estimated system. This model can be used for forecasting and simulation. An alternative approach, creating the model and including the system object by name, is described in Chapter 23, “Specifying a Model”, page 551).

There are other procedures for working with the system:

- **Estimate...** opens the dialog for estimating the system of equations. It may also be accessed by pressing **Estimate** on the system toolbar.
- **Make Residuals** creates a number of series containing the residuals for each equation in the system. The residuals will be given the next unused name of the form RESID01, RESID02, etc., in the order that the equations are specified in the system.
- **Make Endogenous Group** creates an untitled group object containing the endogenous variables.

## Example

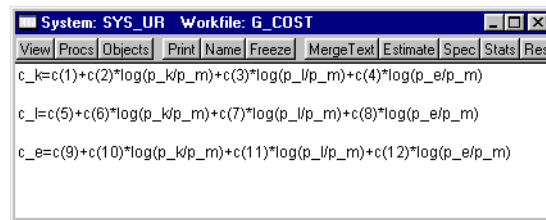
As an illustration of the process of estimating a system of equations in EViews, we estimate a translog cost function using data from Berndt and Wood (1975) as tabulated in Greene (1997). The translog cost function has four factors with three equations of the form:

$$\begin{aligned}
 c_K &= \beta_K + \delta_{KK} \log\left(\frac{p_K}{p_M}\right) + \delta_{KL} \log\left(\frac{p_L}{p_M}\right) + \delta_{KE} \log\left(\frac{p_E}{p_M}\right) + \epsilon_K \\
 c_L &= \beta_L + \delta_{LK} \log\left(\frac{p_K}{p_M}\right) + \delta_{LL} \log\left(\frac{p_L}{p_M}\right) + \delta_{LE} \log\left(\frac{p_E}{p_M}\right) + \epsilon_L \\
 c_E &= \beta_E + \delta_{EK} \log\left(\frac{p_K}{p_M}\right) + \delta_{EL} \log\left(\frac{p_L}{p_M}\right) + \delta_{EE} \log\left(\frac{p_E}{p_M}\right) + \epsilon_E
 \end{aligned} \tag{19.2}$$

where  $c_i$  and  $p_i$  are the cost share and price of factor  $i$ , respectively.  $\beta$  and  $\delta$  are the parameters to estimate. Note that there are cross equation coefficient restrictions that ensure symmetry of the cross partial derivatives.

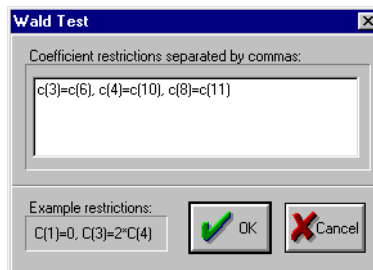
We first estimate this system without imposing the cross equation restrictions and test whether the symmetry restrictions hold. Create a system by clicking **Objects/New Object/System** in the main toolbar or type `system` in the command window. Press the **Name** button and type in the name “SYS\_UR” to name the system.

Next, type in the system window and specify the system as:



We estimate this model by full information maximum likelihood (FIML). FIML is invariant to the equation that is dropped. Press the **Estimate** button and choose **Full Information Maximum Likelihood**. EViews presents the estimated coefficients and regression statistics for each equation.

To test the symmetry restrictions, select **View/Wald Coefficient Tests...**, fill in the dialog as



and click OK. The result:

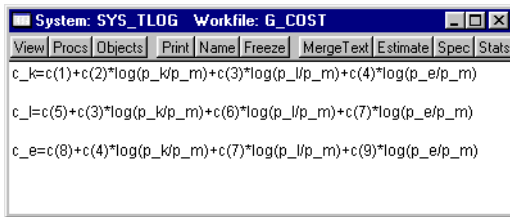
Wald Test:

System: SYS\_UR

Null Hypothesis: C(3)=C(6)  
C(4)=C(10)  
C(8)=C(11)

Chi-square	4.983713	Probability	0.172994
------------	----------	-------------	----------

fails to reject the symmetry restrictions. To estimate the system imposing the symmetry restrictions, click **View/System Specification** and modify the system.



We have named the system SYS\_TLOG. Note that to impose symmetry in the translog specification, we have restricted the coefficients on the cross-price terms to be the same (we have also renumbered the 9 remaining coefficients so that they are consecutive). The restrictions are imposed by using the same coefficients in each equation. For example, the coefficient on the  $\log(P_L/P_M)$  term in the  $C_K$  equation,  $C(3)$ , is the same as the coefficient on the  $\log(P_K/P_M)$  term in the  $C_L$  equation.

To estimate this model using FIML, click **Estimate** and choose **Full Information Maximum Likelihood**. The top part of the equation describes the estimation specification, and provides coefficient and standard error estimates,  $t$ -statistics,  $p$ -values, and summary statistics:

System: SYS\_TLOG

Estimation Method: Full Information Maximum Likelihood (Marquardt)

Date: 08/07/97 Time: 13:17

Sample: 1947 1971

Convergence achieved after 9 iterations

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	0.056672	0.002803	20.21585	0.0000
C(2)	0.029082	0.008310	3.499789	0.0008
C(3)	0.000274	0.009810	0.027925	0.9778
C(4)	-0.010639	0.004931	-2.157489	0.0346
C(5)	0.252923	0.002172	116.4713	0.0000
C(6)	0.076621	0.010717	7.149671	0.0000
C(7)	-0.003500	0.007603	-0.460326	0.6468
C(8)	0.043860	0.002651	16.54573	0.0000
C(9)	0.020018	0.011049	1.811797	0.0746
Log Likelihood		344.4730		
Determinant residual covariance		2.16E-16		

The log likelihood value reported at the bottom of the first part of the table can be used to construct likelihood ratio tests.

Since maximum likelihood assumes the errors are multivariate normal, we may wish to test whether the residuals are normally distributed. Click **Procs/Make Residuals** and EViews opens an untitled group window containing the residuals of each equation in the system. Then to compute descriptive statistics for each residual in the group, select **View/Descriptive Stats** from the group window toolbar:

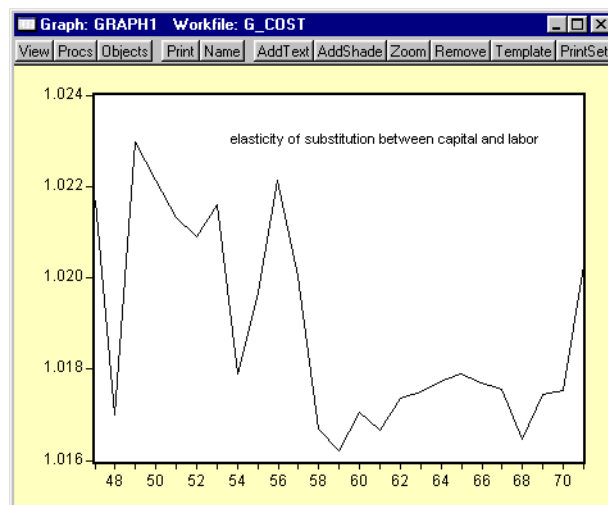
The Jarque-Bera statistic rejects the hypothesis of normal distribution for the second equation but not for the other equations.

The estimated coefficients of the translog cost function can be used to construct estimates of the elasticity of substitution between factors of production. For example, the elasticity of substitution between capital and labor is given by  $1 + c(3) / (C_K * C_L)$ . Note that the elasticity of substitution is not a constant, and depends on the values of  $C_K$  and  $C_L$ . To create a series containing the elasticities computed for each observation, select **Quick/Generate Series...**, and enter:

$$es\_kl = 1 + sys1.c(3) / (c\_k * c\_l)$$

To plot the series of elasticity of substitution between capital and labor for each observation, double click on the series name ES\_KL in the workfile and select **View/Line Graph**:

Group: UNTITLED    Workfile: G_COST				
View	Procs	Objects	Print	Name
Freeze	Sample	Sheet	Stats	Spec
	RESID01	RESID02	RESID03	
Mean	0.000118	0.000168	7.84E-05	
Median	-0.000451	0.000231	-0.000843	
Maximum	0.007941	0.018095	0.002967	
Minimum	-0.005602	-0.011877	-0.002677	
Std. Dev.	0.003233	0.005456	0.001814	
Skewness	0.698243	0.860800	0.130873	
Kurtosis	2.833960	6.206394	1.445040	
Jarque-Bera	2.060150	13.79674	2.590011	
Probability	0.356980	0.001009	0.273896	
Observations	25	25	25	



While it varies over the sample, the elasticity of substitution is generally close to one, which is consistent with the assumption of a Cobb-Douglas cost functions.

## Commands

To create a new system, follow the system command with a name for the system:

```
system demand1
```

creates and opens the window for a new system named DEMAND1.

To estimate a system, follow the name of the system with a dot and the name of the estimation procedure. For example,

```
sys1.fiml
```

estimates SYS1 by full information maximum likelihood.

See the *Command and Programming Reference* for a complete list of commands and options available for system objects.

## Technical Discussion

While the discussion to follow is expressed in terms of a balanced system of linear equations, the analysis carries forward in a straightforward way to unbalanced systems containing nonlinear equations.

Denote a system of  $m$  equations, in stacked form, as

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{bmatrix} = \begin{bmatrix} X_1 & 0 & \dots & 0 \\ 0 & X_2 & & \vdots \\ & & \ddots & 0 \\ 0 & \dots & 0 & X_M \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_M \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_M \end{bmatrix} \quad (19.3)$$

where  $y_m$  is  $T$  vector,  $X_m$  is a  $T \times k_m$  matrix, and  $\beta_m$  is a  $k_m$  vector of coefficients. The error terms  $\epsilon$  have an  $MT \times MT$  covariance matrix  $V$ . The system may be written in compact form as:

$$y = X\beta + \epsilon. \quad (19.4)$$

Under the standard assumptions, the residual variance matrix from this stacked system is given by

$$V = E(\epsilon\epsilon') = \sigma^2(I_M \otimes I_T). \quad (19.5)$$

Other residual structures are of interest. First, the errors may be heteroskedastic across the  $m$  equations. Second, they may be heteroskedastic and contemporaneously correlated. We can

characterize both of these cases by defining the  $M \times M$  matrix of contemporaneous correlations,  $\Sigma$ , where the  $ij$ -th element of  $\Sigma$  is given by  $\sigma_{ij} = E(\epsilon_{it}\epsilon_{jt})$  for all  $t$ . If the errors are contemporaneously uncorrelated, then,  $\sigma_{ij} = 0$  for  $i \neq j$ , and we can write:

$$V = \text{diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_M^2) \otimes I_T \quad (19.6)$$

More generally, if the errors are heteroskedastic and contemporaneously correlated:

$$V = \Sigma \otimes I_T. \quad (19.7)$$

Lastly, at the most general level, there may be heteroskedasticity, contemporaneous correlation, and autocorrelation of the residuals. The general variance matrix of the residuals may be written:

$$V = \begin{pmatrix} \sigma_{11}\Sigma_{11} & \sigma_{12}\Sigma_{12} & \dots & \sigma_{1M}\Sigma_{1M} \\ \sigma_{21}\Sigma_{21} & \sigma_{22}\Sigma_{22} & & \vdots \\ & & \ddots & \\ \sigma_{M1}\Sigma_{M1} & \dots & & \sigma_{MM}\Sigma_{MM} \end{pmatrix} \quad (19.8)$$

where  $\Sigma_{ij}$  is an autocorrelation matrix for the  $i$ -th and  $j$ -th equations.

### Ordinary least squares

The OLS estimator of the estimated variance matrix of the parameters is valid under the assumption that  $V = \Sigma \otimes I_T$ . The estimator for  $\beta$  is given by:

$$b_{LS} = (X'X)^{-1}X'y, \quad (19.9)$$

and the variance estimator is given by

$$\text{var}(b_{LS}) = s^2(X'X)^{-1}, \quad (19.10)$$

where  $s^2$  is the residual variance estimate for the stacked system.

### Weighted least squares

The weighted least squares estimator is given by:

$$b_{WLS} = (X'\hat{V}^{-1}X)^{-1}X'\hat{V}^{-1}y \quad (19.11)$$

where  $\hat{V} = \text{diag}(s_{11}, s_{22}, \dots, s_{MM}) \otimes I_T$  is a consistent estimator of  $V$ , and  $s_{ii}$  is the residual variance estimator,

$$s_{ij} = ((y_i - X_i b_{LS})'(y_j - X_j b_{LS}))/\max(T_i, T_j) \quad (19.12)$$

where the inner product is taken over the non-missing common elements of  $i$  and  $j$ . The max function in Equation (19.12) is designed to handle the case of unbalanced data by down-weighting the covariance terms. Provided the missing values are asymptotically negligible, this yields a

consistent estimator of the variance elements. Note also that there is no adjustment for degrees-of-freedom.

In estimation, are given a choice of which coefficients to use in computing the  $s_{ij}$ . If you choose not to iterate the weights, the OLS coefficient estimates will be used to estimate the variances. If you choose to iterate the weights, the current parameter estimates (which may be based on the previously computed weights) are used in computing the  $s_{ij}$ . This latter procedure may be iterated until the weights and coefficients converge.

The estimator for the coefficient variance matrix is:

$$\text{var}(b_{WLS}) = (X' \hat{V}^{-1} X)^{-1} . \quad (19.13)$$

The weighted least squares estimator is efficient, and the variance estimator consistent, under the assumption that there is heteroskedasticity, but no serial or contemporaneous correlation in the residuals.

It is worth pointing out that if there are no cross-equation restrictions on the parameters of the model, weighted LS on the entire system yields estimates that are identical to those obtained by equation-by-equation LS. Consider the following simple model:

$$\begin{aligned} y_1 &= X_1 \beta_1 + \epsilon_1 \\ y_2 &= X_2 \beta_2 + \epsilon_2 \end{aligned} \quad (19.14)$$

If  $\beta_1$  and  $\beta_2$  are unrestricted, the WLS estimator given in Equation (19.13) yields,

$$b_{WLS} = \begin{bmatrix} ((X_1' X_1)/s_{11})^{-1} ((X_1' y_1)/s_{11}) \\ ((X_2' X_2)/s_{22})^{-1} ((X_2' y_2)/s_{22}) \end{bmatrix} = \begin{bmatrix} (X_1' X_1)^{-1} X_1' y_1 \\ (X_2' X_2)^{-1} X_2' y_2 \end{bmatrix}. \quad (19.15)$$

The expression on the right is equivalent to equation-by-equation OLS. Note, however, that even without cross-equation restrictions, the standard errors are not the same in the two cases.

### Seemingly unrelated regression (SUR)

SUR is appropriate when all the right-hand side regressors  $X$  are assumed to be exogenous, and the errors are heteroskedastic and contemporaneously correlated so that the error variance matrix is given by  $V = \Sigma \otimes I_T$ . Zellner's SUR estimator of  $\beta$  takes the form

$$b_{SUR} = (X' (\hat{\Sigma} \otimes I_T)^{-1} X)^{-1} X' (\hat{\Sigma} \otimes I_T)^{-1} y, \quad (19.16)$$

where  $\hat{\Sigma}$  is a consistent estimate of  $\Sigma$  with typical element  $s_{ij}$ , for all  $i$  and  $j$ .

If you include AR terms in equation  $j$ , EViews transforms the model (see Chapter 13, "Estimating AR Models" on page 277, for details) and estimates the following equation:



$$y_{jt} = X_{jt}\beta_j + \left( \sum_{r=1}^{p_j} \rho_{jr}(y_{j(t-r)} - X_{j(t-r)}) \right) + \epsilon_{jt} \quad (19.17)$$

where  $\epsilon_j$  is assumed to be serially independent, but possibly correlated contemporaneously across equations. At the beginning of the first iteration, we estimate the equation by nonlinear LS and use the estimates to compute the residuals  $\hat{\epsilon}$ . We then construct an estimate of  $\Sigma$  using  $s_{ij} = (\hat{\epsilon}_i \hat{\epsilon}_j) / \max(T_i, T_j)$  and perform nonlinear GLS to complete one iteration of the estimation procedure. These iterations are repeated until the coefficients and weights converge.

### Two stage least squares (TSLS) and Weighted TSLS

TSLS is a single equation estimation method that is appropriate when some of the variables in  $X$  are endogenous. Write the  $j$ -th equation of the system as

$$Y\Gamma_j + XB_j + \epsilon_j = 0 \quad (19.18)$$

or, alternatively,

$$y_j = Y_j\gamma_j + X_j\beta_j + \epsilon_j = Z_j\delta_j + \epsilon_j \quad (19.19)$$

where  $\Gamma_j' = (-1, \gamma_j', 0)$ ,  $B_j' = (\beta_j', 0)$ ,  $Z_j' = (Y_j', X_j')$  and  $\delta_j' = (\gamma_j', \beta_j')$ .  $Y$  is the matrix of endogenous variables and  $X$  is the matrix of exogenous variables.

In the first stage, we regress the right-hand side endogenous variables  $y_j$  on all exogenous variables  $X$  and get the fitted values

$$\hat{y}_j = X(X'X)^{-1}X'y_j. \quad (19.20)$$

In the second stage, we regress  $y_j$  on  $\hat{Y}_j$  and  $X_j$  to get

$$\hat{\delta}_{2SLS} = (\hat{Z}_j' \hat{Z}_j)^{-1} \hat{Z}_j' y_j. \quad (19.21)$$

where  $\hat{Z}_j = (\hat{Y}_j, X_j)$ .

Weighted TSLS applies the weights in the second stage so that

$$\hat{\delta}_{W2SLS} = (\hat{Z}_j' \hat{V}^{-1} \hat{Z}_j)^{-1} \hat{Z}_j' \hat{V}^{-1} y_j \quad (19.22)$$

where the elements of the variance matrix are estimated in the usual fashion using the residuals from unweighted TSLS.

If you choose to iterate the weights,  $\Sigma$  is estimated at each step using the current values of the coefficients and residuals.

### Three stage least squares (3SLS)

Since TSLS is a single equation estimator that does not take account of the covariances between residuals, it is not, in general, fully efficient. 3SLS is a system method that estimates all of the coefficients of the model, then forms weights and reestimates the model using the estimated

weighting matrix. It should be viewed as the endogenous variable analogue to the SUR estimator described above,

The first two stages of 3SLS are the same as in TSLS. In the third stage, we apply feasible generalized least squares (FGLS) to the equations in the system in a manner analogous to the SUR estimator.

SUR uses the OLS residuals to obtain a consistent estimate of the cross-equation covariance matrix  $\Sigma$ . This covariance estimator is not, however, consistent if any of the right-hand side variables are endogenous. 3SLS uses the 2SLS residuals to obtain a consistent estimate of  $\Sigma$ .

$$\hat{\delta}_{3SLS} = (Z(\hat{\Sigma}^{-1} \otimes X(X'X)^{-1}X')Z)^{-1}Z(\hat{\Sigma}^{-1} \otimes X(X'X)^{-1}X')y, \quad (19.23)$$

where  $\hat{\Sigma}$  has typical element

$$s_{ij} = ((y_i - Z_i\hat{\gamma}_{2SLS})'(y_j - Z_j\hat{\gamma}_{2SLS}))/\max(T_i, T_j). \quad (19.24)$$

If you choose to iterate the weights, the current coefficients and residuals will be used to for the estimate  $\hat{\Sigma}$ .

### Generalized Method of Moments (GMM)

The basic idea underlying GMM is simple and intuitive. We have a set of theoretical moment conditions that the parameters of interest  $\theta$  should satisfy. We denote these moment conditions as

$$E(m(y, \theta)) = 0. \quad (19.25)$$

The method of moments estimator is defined by replacing the moment condition (19.25) by its sample analog

$$\left(\sum_t m(y_t, \theta)\right)/T = 0. \quad (19.26)$$

However, condition (2) will not be satisfied for any  $\theta$  when there are more restrictions  $m$  than there are parameters  $\theta$ . To allow for such overidentification, the GMM estimator is defined by minimizing the following criterion function:

$$\sum_t m(y_t, \theta)' A(y_t, \theta) m(y_t, \theta), \quad (19.27)$$

which measures the “distance” between  $m$  and zero.  $A$  is a weighting matrix that weights each moment condition. Any symmetric positive definite matrix  $A$  will yield a consistent estimate of  $\theta$ . However, it can be shown that a necessary (but not sufficient) condition to obtain an (asymptotically) efficient estimate of  $\theta$  is to set  $A$  equal to the inverse of the covariance matrix  $\Omega$  of the sample moments  $m$ . This follows intuitively, since we want to put less weight on the conditions that are more imprecise.

To obtain GMM estimates in EViews, you must be able to write the moment conditions in Equation (19.25) as an orthogonality condition between the residuals of a regression equation,  $u(y, \theta, X)$ , and a set of instrumental variables,  $Z$ , so that

$$m(\theta, y, X, Z) = Z'u(\theta, y, X) \quad (19.28)$$

For example, the OLS estimator is obtained as a GMM estimator with the orthogonality conditions

$$X'(y - X\beta) = 0. \quad (19.29)$$

For the GMM estimator to be identified, there must be at least as many instrumental variables  $Z$  as there are parameters  $\theta$ . See the section on “Generalized Method of Moments (GMM)” in Chapter 12 (beginning on page 267), for additional examples of GMM orthogonality conditions.

An important aspect of specifying a GMM problem is the choice of the weighting matrix  $A$ . EViews uses the optimal  $A = \hat{\Omega}^{-1}$ , where  $\hat{\Omega}$  is the estimated covariance matrix of the sample moments  $m$ . EViews uses the consistent TSLS estimates for the initial estimate of  $\theta$  in forming the estimate of  $\Omega$ .

### White's Heteroskedasticity Consistent Covariance Matrix

If you choose the **GMM-Cross section** option, EViews estimates  $\Omega$  using White's heteroskedasticity consistent covariance matrix

$$\hat{\Omega}_W = \hat{I}(0) = \frac{1}{T-k} \left( \sum_{t=1}^T Z_t' u_t u_t' Z_t \right) \quad (19.30)$$

where  $u$  is the vector of residuals, and  $Z_t$  is a  $k \times p$  matrix such that the  $p$  moment conditions at  $t$  may be written as  $m(\theta, y_t, X_t, Z_t) = Z_t' u(\theta, y_t, X_t)$ .

### Heteroskedasticity and Autocorrelation Consistent (HAC) Covariance Matrix

If you choose the **GMM-Time series** option, EViews estimates  $\Omega$  by

$$\hat{\Omega}_{HAC} = \hat{I}(0) + \left( \sum_{j=1}^{T-1} k(j, q) (\hat{I}(j) - \hat{I}'(j)) \right) \quad (19.31)$$

where

$$\hat{I}(j) = \frac{1}{T-k} \left( \sum_{t=j+1}^T Z_{t-j}' u_t u_{t-j}' Z_t \right).$$

You also need to specify the *kernel*  $\kappa$  and the *bandwidth*  $q$ .

### Kernel Options

The kernel  $\kappa$  is used to weight the covariances so that  $\hat{\Omega}$  is ensured to be positive semi-definite. EViews provides two choices for the kernel, Bartlett and quadratic spectral (QS). The Bartlett kernel is given by

$$\kappa(j, q) = \begin{cases} 1 - \frac{j}{q} & 0 \leq j \leq q \\ 0 & \text{otherwise} \end{cases} \quad (19.32)$$

while the quadratic spectral (QS) kernel is given by

$$k(j/q) = \frac{25}{12(\pi x)^2} \left( \frac{\sin(6\pi x/5)}{6\pi x/5} - \cos(6\pi x/5) \right) \quad (19.33)$$

where  $x = j/q$ . The QS has a faster rate of convergence than the Bartlett and is smooth and not truncated (Andrews 1991). Note that even though the QS kernel is not truncated, it still depends on the bandwidth  $q$  (which need not be an integer).

### Bandwidth Selection

The bandwidth  $q$  determines how the weights given by the kernel change with the lags in the estimation of  $\Omega$ . Newey-West fixed bandwidth is based solely on the number of observations in the sample and is given by

$$q = \text{int}(4(T/100)^{2/9}) \quad (19.34)$$

where  $\text{int}(\cdot)$  denotes the integer part of the argument.

EViews also provides two “automatic”, or data dependent bandwidth selection methods that are based on the autocorrelations in the data. Both methods select the bandwidth according to

$$q = \begin{cases} \text{int}(1.1447(\hat{\alpha}(1)T)^{1/3}) & \text{for the Bartlett kernel} \\ 1.3221(\hat{\alpha}(2)T)^{1/5} & \text{for the QS kernel} \end{cases} \quad (19.35)$$

The two methods, Andrews and Variable-Newey-West, differ in how they estimate  $\hat{\alpha}(1)$  and  $\hat{\alpha}(2)$ .

Andrews (1991) is a parametric method that assumes the sample moments follow an AR(1) process. We first fit an AR(1) to each sample moment (19.28) and estimate the autocorrelation coefficients  $\hat{\rho}_i$  and the residual variances  $\hat{\sigma}_i^2$  for  $i = 1, 2, \dots, zn$ , where  $z$  is the number of instrumental variables and  $n$  is the number of equations in the system. Then  $\hat{\alpha}(1)$  and  $\hat{\alpha}(2)$  are estimated by

$$\begin{aligned} \hat{\alpha}(1) &= \left( \sum_{i=1}^{zn} \frac{4\hat{\rho}_i^2 \hat{\sigma}_i^4}{(1-\hat{\rho}_i)^6 (1+\hat{\rho}_i)^2} \right) / \left( \sum_{i=1}^{zn} \frac{\hat{\sigma}_i^4}{(1-\hat{\rho}_i)^4} \right) \\ \hat{\alpha}(2) &= \left( \sum_{i=1}^{zn} \frac{4\hat{\rho}_i^2 \hat{\sigma}_i^4}{(1-\hat{\rho}_i)^8} \right) / \left( \sum_{i=1}^{zn} \frac{\hat{\sigma}_i^4}{(1-\hat{\rho}_i)^4} \right) \end{aligned} \quad (19.36)$$

Note that we weight all moments equally, including the moment corresponding to the constant.

Newey-West (1994) is a nonparametric method based on a truncated weighted sum of the estimated cross-moments  $\hat{I}(j)$ .  $\hat{\alpha}(1)$  and  $\hat{\alpha}(2)$  are estimated by

$$\hat{\alpha}(p) = \left( \frac{l' F(p) l}{l' F(0) l} \right) \quad (19.37)$$

where  $l$  is a vector of ones and  $F(p) = \Gamma(0) + \sum_{i=1}^L i^p (\hat{I}(i) + \hat{I}'(i))$ , for  $p = 1, 2$ .

One practical problem with the Newey-West method is that we have to choose a lag selection parameter  $L$ . The choice of  $L$  is arbitrary, subject to the condition that it grow at a certain rate. EViews sets the lag parameter to

$$L = \begin{cases} \text{int}(4(T/100)^{2/9}) & \text{for the Bartlett kernel} \\ T & \text{for the QS kernel} \end{cases} \quad (19.38)$$

## Prewhitening

You can also choose to prewhiten the sample moments  $m$  to “soak up” the correlations in  $m$  prior to GMM estimation. We first fit a VAR(1) to the sample moments

$$m_t = A m_{t-1} + v_t. \quad (19.39)$$

Then the variance of  $\hat{\Omega}$  of  $m$  is estimated by  $\hat{\Omega} = (I - A)^{-1} \hat{\Omega}^* (I - A)^{-1}$  where  $\hat{\Omega}^*$  is the variance of the residuals  $v_t$  and is computed using any of the above methods. The GMM estimator is then found by minimizing the criterion function:

$$u' Z \hat{\Omega}^{-1} Z' u \quad (19.40)$$

Note that while Andrews and Monahan (1992) adjust the VAR estimates to avoid singularity when the moments are near unit root processes, EViews does not perform this eigenvalue adjustment.



## Chapter 20. Vector Autoregression and Error Correction Models

---

The structural approach to simultaneous equations modeling uses economic theory to describe the relationships between several variables of interest. The resulting model is then estimated, and used to test the empirical relevance of the theory.

Unfortunately, economic theory is often not rich enough to provide a tight specification of the dynamic relationship among variables. Furthermore, estimation and inference are complicated by the fact that endogenous variables may appear on both the left and right sides of equations.

These problems lead to alternative, non-structural, approaches to modeling the relationship between several variables. This chapter describes the estimation and analysis of vector autoregression (VAR) and the vector error correction (VEC) models. We also describe tools for testing for the presence of cointegrating relationships among several variables.

### Vector Autoregression Theory

The vector autoregression (VAR) is commonly used for forecasting systems of interrelated time series and for analyzing the dynamic impact of random disturbances on the system of variables.

The VAR approach sidesteps the need for structural modeling by modeling every endogenous variable in the system as a function of the lagged values of *all* of the endogenous variables in the system.

The mathematical form of a VAR is

$$y_t = A_1 y_{t-1} + \dots + A_p y_{t-p} + B x_t + \epsilon_t \quad (20.1)$$

where  $y_t$  is a  $k$  vector of endogenous variables,  $x_t$  is a  $d$  vector of exogenous variables,  $A_1, \dots, A_p$  and  $B$  are matrices of coefficients to be estimated, and  $\epsilon_t$  is a vector of innovations that may be contemporaneously correlated with each other but are uncorrelated with their own lagged values and uncorrelated with all of the right-hand side variables.

Since only lagged values of the endogenous variables appear on the right-hand side of each equation, there is no issue of simultaneity, and OLS is the appropriate estimation technique. Note that the assumption that the disturbances are not serially correlated is

not restrictive because any serial correlation could be absorbed by adding more lagged  $y$ 's.

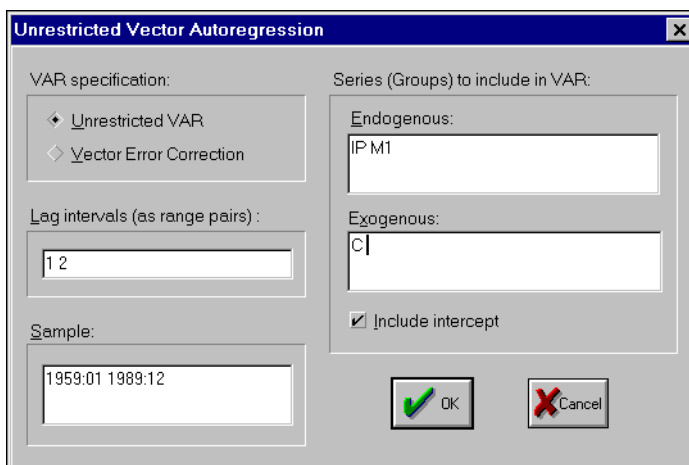
As an example of a VAR, suppose that industrial production (IP) and money supply (M1) are jointly determined by a two variable VAR and let a constant be the only exogenous variable. With two lagged values of the endogenous variables, the VAR is

$$\begin{aligned} IP_t &= a_{11}IP_{t-1} + a_{12}M1_{t-1} + b_{11}IP_{t-2} + b_{12}M1_{t-2} + c_1 + \epsilon_{1t} \\ M1_t &= a_{21}IP_{t-1} + a_{22}M1_{t-1} + b_{21}IP_{t-2} + b_{22}M1_{t-2} + c_2 + \epsilon_{2t} \end{aligned} \quad (20.2)$$

where  $a$ ,  $b$ ,  $c$  are the parameters to be estimated.

## Estimating a VAR in EViews

To specify a vector autoregression, you must create a VAR object. Select **Quick/Estimate VAR...** or type `var` in the command window. The following dialog box should appear:



Fill out the dialog with the appropriate information:

- Enter the lag information in the first edit box. This information tells EViews which lags should be included on the right-hand side of each equation. This information is entered in pairs: each pair of numbers defines a *range* of lags.

For example, the lag pair:

1 2

tells EViews to use the first and second lags of all of the variables in the system as right-hand side variables.

You can add any number of lag intervals, all entered in pairs. The lag specification:

2 4 6 9 11 11



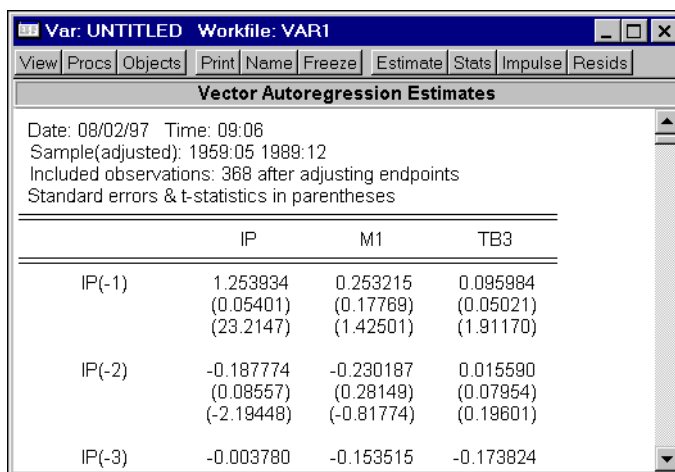
uses lags two through four, lags six through nine, and lag eleven.

- Enter the endogenous and exogenous variables in the appropriate edit boxes.
- Select the specification type: **Unrestricted VAR** or **Vector Error Correction (VEC)**. What we have been calling a VAR is the *unrestricted* VAR. VECs will be explained in detail, below.
- Check or uncheck the box to include or exclude the constant from the right-hand side of the equations.

The example dialog corresponds to the VAR example above: IP and M1 are the endogenous variables, the model uses lags 1 through 2, and the constant is the only exogenous variable. You will note that we have both entered C as an exogenous variable, and checked the box instructing EViews to include a constant in the VAR. This is redundant, but harmless, since EViews will only include a single constant.

## VAR Estimation Output

Once you have specified the VAR, click **OK**. EViews will display the estimation results in the VAR window:



	IP	M1	TB3
IP(-1)	1.253934 (0.05401) (23.2147)	0.253215 (0.17769) (1.42501)	0.095984 (0.05021) (1.91170)
IP(-2)	-0.187774 (0.08557) (-2.19448)	-0.230187 (0.28149) (-0.81774)	0.015590 (0.07954) (0.19601)
IP(-3)	-0.003780	-0.153515	-0.173824

Each column in the table corresponds to the equation for one endogenous variable in the VAR. For each right-hand side variable, EViews reports three numbers: a coefficient point estimate, the estimated coefficient standard error, and the *t*-statistic.

Two types of regression statistics are reported at the bottom of the VAR estimation output.

R-squared	0.999221	0.999915	0.968018
Adj. R-squared	0.999195	0.999912	0.966937
Sum sq. resids	113.8813	1232.453	98.39849
S.E. equation	0.566385	1.863249	0.526478
Log likelihood	-306.3509	-744.5662	-279.4628
Akaike AIC	-1.102274	1.279331	-1.248405
Schwarz SC	-0.964217	1.417388	-1.110348
Mean dependent	70.97919	339.7451	6.333891
S.D. dependent	19.95932	198.6301	2.895381
Determinant Residual Covariance		0.259637	
Log Likelihood		-1318.390	
Akaike Information Criteria		-1.136514	
Schwarz Criteria		-0.722342	

The first part of the output displays the standard regression statistics for each equation. The results are computed separately, using the residuals for each equation and arranged in the appropriate column.

The numbers at the very bottom of the table are the regression statistics for the VAR system. The determinant of the residual covariance is computed as

$$|\hat{\Omega}| = \det\left(\sum_t \tilde{\epsilon}_t \tilde{\epsilon}_t' / T\right) \quad (20.3)$$

where  $\epsilon_t$  is the  $k$  vector of residuals. The log-likelihood value is computed assuming a multivariate normal (Gaussian) distribution as:

$$l = -\frac{Tk}{2}(1 + \log(2\pi)) - \frac{T}{2}\log|\hat{\Omega}| \quad (20.4)$$

and the two information criteria are computed as  $AIC = -2l/T + 2n/T$ , and  $SC = -2l/T + (n\log T)/T$ , where  $n = k(d + pk)$  is the total number of estimated parameters in the VAR. These information criteria can be used for model selection such as determining the lag length of the VAR; the smaller the value of the information criteria, the “better” the model. It is worth noting that by ignoring constant terms, some reference sources define the AIC for a multivariate normal system in a slightly different fashion, as:

$$\log|\hat{\Omega}| + 2n/T. \quad (20.5)$$

See also Appendix F for additional discussion of various information criteria.

## Working with a VAR

Once you have estimated your VAR, EViews provides various views and procedures that allow you to use the estimated VAR for further analyses. In this section, we discuss views that are specific to VARs; for other views and procedures, see the discussion of system views in Chapter 19.

In empirical applications, the main uses of the VAR are the impulse response analysis, variance decomposition, and Granger causality tests.

## Impulse Response Functions

An *impulse response function* traces the effect of a one standard deviation shock to one of the innovations on current and future values of the endogenous variables.

A shock to the  $i$ -th variable directly affects the  $i$ -th variable, and is also transmitted to all of the endogenous variables through the dynamic structure of the VAR.

Consider a simple bivariate VAR(1):

$$\begin{aligned} IP_t &= a_{11}IP_{t-1} + a_{12}M1_{t-1} + \epsilon_{1t} \\ M1_t &= a_{21}IP_{t-1} + a_{22}M1_{t-1} + \epsilon_{2t} \end{aligned} \quad (20.6)$$

A change in  $\epsilon_{1t}$  will immediately change the value of current IP. It will also change all future values of IP and M1 since lagged IP appears in both equations.

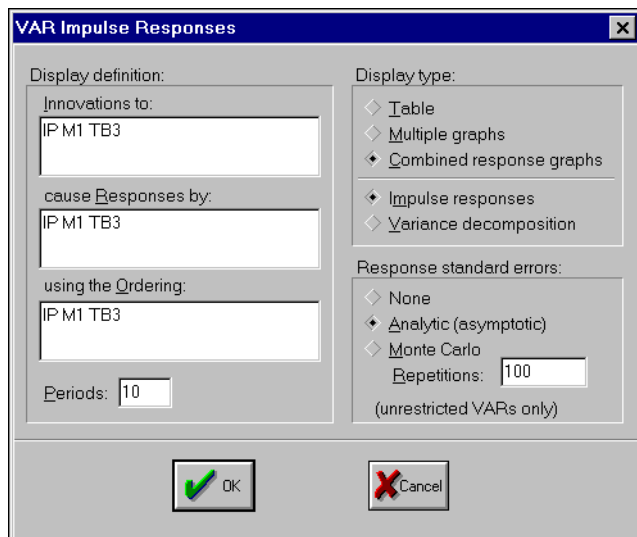
If the innovations,  $\epsilon_{1t}$  and  $\epsilon_{2t}$  in our example, are uncorrelated, interpretation of the impulse response is straightforward.  $\epsilon_{1t}$  is the innovation for IP and  $\epsilon_{2t}$  is the innovation for M1. The impulse response functions for  $\epsilon_{2t}$  measures the effect of a one standard deviation monetary shock on current and future industrial production and money stock.

The innovations are, however, usually correlated, so that they have a common component which cannot be associated with a specific variable. A somewhat arbitrary but common method of dealing with this issue is to attribute all of the effect of any common component to the variable that comes first in the VAR system. In our example, the common component of  $\epsilon_{1t}$  and  $\epsilon_{2t}$  is totally attributed to  $\epsilon_{1t}$ , because  $\epsilon_{1t}$  precedes  $\epsilon_{2t}$ .  $\epsilon_{1t}$  is then the IP innovation, and  $\epsilon_{2t}$ , the M1 innovation, is transformed to remove the common component.

More technically, the errors are orthogonalized by a Cholesky decomposition so that the covariance matrix of the resulting innovations is diagonal—see “Technical Discussion” on page 502 below for details. While the Cholesky decomposition is widely used, it is a rather arbitrary method of attributing common effects. You should be aware that changing the order of equations can dramatically change the impulse responses.

## Generating Impulse Response Functions from VARs

To obtain the impulse response functions for your VAR, select **Impulse** on the VAR toolbar.



• In the top two edit boxes, you should enter the variables for which you wish to generate innovations, and the variables for which you wish to observe the impulse responses. EViews will compute one impulse response function for each innovation and endogenous variable pair. For example, a four variable VAR has 16 potential impulse response functions. The order in which you enter these variables only affects the display of results.

• In the bottom edit box, you should specify the ordering of the variables in the VAR—as explained above, there are significant implications of this choice.

- Next, specify the number of periods for which you wish to trace the response function.
- On the right, specify the form of the output, either a table of numbers, separate graphs of each impulse response function, or graphs that compare the responses of each variable to each of the innovations.
- Choose between impulse responses and variance decomposition (see “Variance Decomposition” on page 503, below).
- Finally, you should make a choice about the standard errors of the response functions. You can choose not to calculate standard errors, to calculate them from the asymptotic analytical formula, or to calculate them by Monte Carlo methods; see “Technical Discussion” on page 502 for further information of these options. For the Monte Carlo option, you should also specify the number of repetitions.

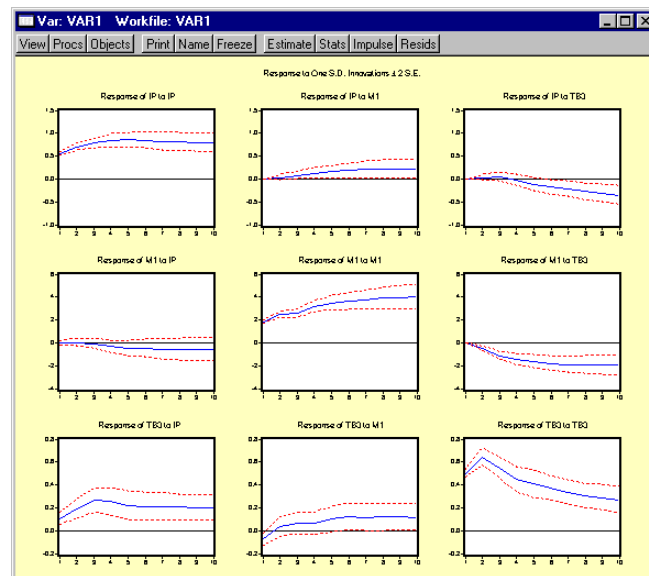
The **Table** option displays the impulse response functions.

Here we see the response of IP to innovations in IP, M1 and TB3. There is a separate table for each endogenous variable which tabulates the responses to all the innovations that you specify. The numbers in parentheses are the standard errors, computed according to your choice of the **Response standard errors** option.

The **Combined response graphs** option displays these tabular results in a set of graphs, with one graph corresponding to each endogenous variable table.

The **Multiple graphs** option displays the impulse response functions with a separate graph for each endogenous variable and response pair:

Var: VAR1 Workfile: VAR1			
View Procs Objects Print Name Freeze Estimate Stat			
Impulse Response to One S.D. Innovations			
Response of IP:			
Period	IP	M1	TB3
1	0.556291 (0.02051)	0.000000 (0.00000)	0.000000 (0.00000)
2	0.703486 (0.03889)	0.030134 (0.02751)	0.029316 (0.02881)
3	0.784828 (0.05491)	0.072117 (0.04310)	0.044857 (0.04633)
4	0.841694	0.129281	-0.027976



If you choose to compute the standard errors, EViews displays the plus/minus two standard deviation bands, alongside the impulse responses. The standard error bands will not be displayed if you select the **Combined response graph** option.

## Variance Decomposition

Variance decomposition provides a different method of depicting the system dynamics. Impulse response functions trace the effects of a shock to an endogenous variable on the variables in the

VAR. By contrast, variance decomposition decomposes variation in an endogenous variable into the component shocks to the endogenous variables in the VAR. The variance decomposition gives information about the relative importance of each random innovation to the variables in the VAR. See “Technical Discussion” on page 502 for additional details.

### How to compute variance decomposition from VARs

To obtain the variance decomposition of a VAR, click **Impulse** in the VAR toolbar and choose the **Variance decomposition** option. You should provide the same information as for impulse responses above, except that the choice of innovations is not needed. Note that there are no standard errors for variance decomposition.

The **Table** option displays the variance decomposition in tabular form. The first few lines of the table for our example are given by:

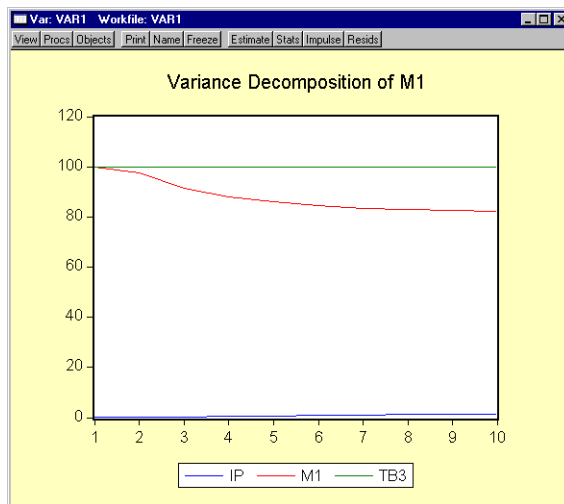
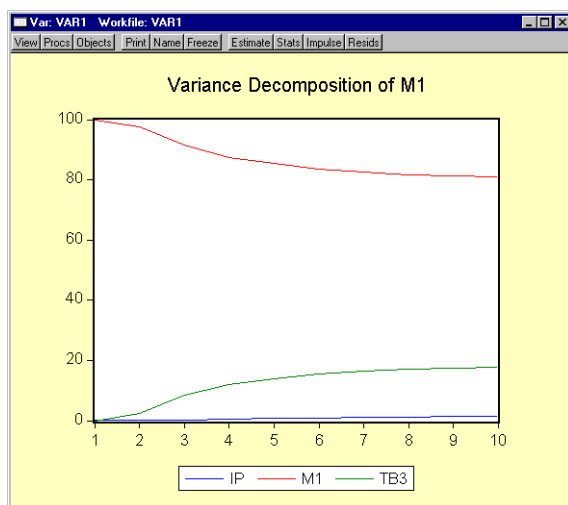
Variance Decomposition of IP:				
Period	S.E.	IP	M1	TB3
1	0.556291	100.0000	0.000000	0.000000
2	0.897842	99.78074	0.112644	0.106613
3	1.195528	99.37168	0.427407	0.200912
4	1.468071	98.77151	1.058935	0.169553
5	1.708867	97.79550	1.655343	0.549154
6	1.920977	96.50807	2.228007	1.263924
7	2.113696	94.98790	2.801220	2.210878
8	2.292034	93.37450	3.283985	3.341518
9	2.458283	91.73909	3.668923	4.591985
10	2.614520	90.11416	3.983015	5.902828

Variance Decomposition of M1:				
Period	S.E.	IP	M1	TB3
1	1.830043	0.009722	99.99028	0.000000
2	3.107362	0.004666	97.44806	2.547270
3	4.199771	0.043461	91.39424	8.562295
4	5.458252	0.401417	87.48964	12.10894
5	6.697880	0.722382	85.36339	13.91423

EViews displays a separate variance decomposition for each endogenous variable. The column S.E. is the forecast error of the variable for each forecast horizon. The source of this forecast error is the variation in current and future values of the innovations to each endogenous variable in the VAR.

The remaining columns give the percentage of the variance due to each innovation; each row adds up to 100. For the variable that comes first in the VAR ordering, the only source of the one period ahead variation is its own innovation and the first number is always 100 percent. We emphasize again that the variance decomposition depends critically on the ordering of equations. The **Combined response graphs** option plots the decomposition of each forecast variance as line graphs.



The top graph shows the EViews default: the variance decomposition is displayed as separate line graphs with the  $y$ -axis height measuring the relative importance of each innovation. You can also change the line graphs to stacked lines; double click in the white area of the graph and change the **Graph Type** to **Stacked Lines**. The stacked line graph (the bottom graph) shows the relative importance of each innovation as the height between each successive line. Note that the last line, corresponding to the innovation that comes last in the VAR ordering, is always flat at 100 because of the adding up constraint.

## Granger Causality Tests

You can test Granger causality by running a VAR on the system of equations and testing for zero restrictions on the VAR coefficients (see Chapter 8, “Granger Causality” on page 193, for a discussion of Granger’s definition of causality). We recommend, however, that you *not* use a VAR to test for Granger Causality, since EViews provides alternative tools that are either easier to use, or more generally applicable than the standard VAR.

You should perform your Granger Causality test in EViews using a group or a system object. Your choice of approach should depend upon your particular testing situation:

Displaying the Granger test view of a group object is the easiest way to perform the test (page 193). The primary limitation of this approach is that in constructing the test statistic, you are restricted to using VARs with a constant as the only exogenous variable.

The system object allows you to include additional exogenous variables in the test, and to impose cross-equation restrictions on the coefficients of the VAR. It is, however, more difficult to set up the test using this approach than using the group Granger test view.

Note that if you wish, you could perform the Granger test using single equation methods. For example, suppose you have a VAR(4) of M1, IP, TB3 with a constant and linear trend and you want to test whether the lagged M1’s are jointly significant in the IP equation. One way to test this hypothesis is to estimate the IP equation using single equation least squares:

```
ls ip m1(-1 to -4) ip(-1 to -4) tb3(-1 to -4) c @trend(50.1)
```

and from the equation window choose **View/Coefficient Tests/Wald** and type the restrictions in the Wald Test dialog as

```
c(1)=0, c(2)=0, c(3)=0, c(4)=0
```

Note that the estimated OLS equation should be identical to the equation estimated by estimating the VAR.

## Obtaining the Coefficients of a VAR

Estimated coefficients of VARs can be accessed by reference to a two dimensional array C. To see the correspondence between each element of C and the estimated coefficients, select **View/Representations** from the VAR toolbar that looks as follows:



Estimation Command:

=====

LS 1 4 IP M1 TB3 @ C

VAR Model:

=====

$IP = C(1,1)*IP(-1) + C(1,2)*IP(-2) + C(1,3)*IP(-3) + C(1,4)*IP(-4) + C(1,5)*M1(-1) + C(1,6)*M1(-2) + C(1,7)*M1(-3) + C(1,8)*M1(-4) + C(1,9)*TB3(-1) + C(1,10)*TB3(-2) + C(1,11)*TB3(-3) + C(1,12)*TB3(-4) + C(1,13)$

$M1 = C(2,1)*IP(-1) + C(2,2)*IP(-2) + C(2,3)*IP(-3) + C(2,4)*IP(-4) + C(2,5)*M1(-1) + C(2,6)*M1(-2) + C(2,7)*M1(-3) + C(2,8)*M1(-4) + C(2,9)*TB3(-1) + C(2,10)*TB3(-2) + C(2,11)*TB3(-3) + C(2,12)*TB3(-4) + C(2,13)$

$TB3 = C(3,1)*IP(-1) + C(3,2)*IP(-2) + C(3,3)*IP(-3) + C(3,4)*IP(-4) + C(3,5)*M1(-1) + C(3,6)*M1(-2) + C(3,7)*M1(-3) + C(3,8)*M1(-4) + C(3,9)*TB3(-1) + C(3,10)*TB3(-2) + C(3,11)*TB3(-3) + C(3,12)*TB3(-4) + C(3,13)$

The first number in C refers to the equation number of the VAR, while the second number in C refers to the variable number in each equation. For example, C(2,3) is the coefficient of the third regressor in the second equation of the VAR. The C(2,3) coefficient of a VAR named VAR01 can then be accessed by the command

```
var01.c(2,3)
```

## Vector Error Correction and Cointegration Theory

The finding that many macro time series may contain a unit root has spurred the development of the theory of non-stationary time series analysis. Engle and Granger (1987) pointed out that a linear combination of two or more non-stationary series may be stationary. If such a stationary, or  $I(0)$ , linear combination exists, the non-stationary (with a unit root), time series are said to be *cointegrated*. The stationary linear combination is called the *cointegrating equation* and may be interpreted as a long-run equilibrium relationship between the variables. For example, consumption and income are likely to be cointegrated. If they were not, then in the long-run consumption might drift above or below income, so that consumers were irrationally spending or piling up savings.

A vector error correction (VEC) model is a restricted VAR that has cointegration restrictions built into the specification, so that it is designed for use with nonstationary series that are known to be cointegrated. The VEC specification restricts the long-run behavior of the endogenous variables to converge to their cointegrating relationships while allowing a wide range of short-run dynamics. The cointegration term is known as the *error correction* term since the deviation from long-run equilibrium is corrected gradually through a series of partial short-run adjustments.

As a simple example, consider a two variable system with one cointegrating equation and no lagged difference terms. The cointegrating equation is

$$y_{2t} = \beta y_{1t} \quad (20.7)$$

and the VEC is

$$\begin{aligned} \Delta y_{1t} &= \gamma_1(y_{2(t-1)} - \beta y_{1(t-1)}) + \epsilon_{1t} \\ \Delta y_{2t} &= \gamma_2(y_{2(t-1)} - \beta y_{1(t-1)}) + \epsilon_{2t} \end{aligned} \quad (20.8)$$

In this simple model, the only right-hand side variable is the error correction term. In long run equilibrium, this term is zero. However, if  $y_1$  and  $y_2$  deviated from long run equilibrium last period, the error correction term is nonzero and each variable adjusts to partially restore the equilibrium relation. The coefficients  $\gamma_1$  and  $\gamma_2$  measure the speed of adjustment.

In this model, the two endogenous variables  $y_{1t}$  and  $y_{2t}$  will have nonzero means but the cointegrating equation will have a zero intercept. To keep the example simple, despite the fact that the use of lagged differences is common, we have included no lagged differences on the right-hand side.

If the two endogenous variables  $y_{1t}$  and  $y_{2t}$  have no trend and the cointegrating equations have an intercept, the VEC has the form

$$\begin{aligned} \Delta y_{1t} &= \gamma_1(y_{2(t-1)} - \mu - \beta y_{1(t-1)}) + \epsilon_{1t} \\ \Delta y_{2t} &= \gamma_2(y_{2(t-1)} - \mu - \beta y_{1(t-1)}) + \epsilon_{2t} \end{aligned} \quad (20.9)$$

Another VEC specification assumes that there are linear trends in the series and a constant in the cointegrating equations, so that it has the form

$$\begin{aligned} \Delta y_{1t} &= \delta_1 + \gamma_1(y_{2(t-1)} - \mu - \beta y_{1(t-1)}) + \epsilon_{1t} \\ \Delta y_{2t} &= \delta_2 + \gamma_2(y_{2(t-1)} - \mu - \beta y_{1(t-1)}) + \epsilon_{2t} \end{aligned} \quad (20.10)$$

Similarly, there may be a trend in the cointegrating equation, but no separate trends in the two VEC equations. Lastly, if there is a separate linear trend outside the parentheses in each VEC equation, then there is an implicit quadratic trend in the series.

For additional discussion of VAR and VEC models, see Davidson and MacKinnon (1993, pp. 715–730), and Hamilton (1994, Chapter 19, pp. 571–629).

## Testing for Cointegration

Given a group of non-stationary series, we may be interested in determining whether the series are cointegrated, and if they are, in identifying the cointegrating (long-run equilibrium) relationships. EViews implements VAR- based cointegration tests using the methodology developed by Johansen (1991, 1995). Johansen's method is to test the restrictions imposed by cointegration on the unrestricted VAR involving the series.

## Johansen's Cointegration Test

Consider a VAR of order  $p$ :

$$y_t = A_1 y_{t-1} + \dots + A_p y_{t-p} + Bx_t + \epsilon_t, \quad (20.11)$$

where  $y_t$  is a  $k$ -vector of non-stationary,  $I(1)$  variables,  $x_t$  is a  $d$  vector of deterministic variables, and  $\epsilon_t$  is a vector of innovations. We can rewrite the VAR as:

$$\Delta y_t = \Pi y_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} + Bx_t + \epsilon_t, \quad (20.12)$$

where

$$\Pi = \left( \sum_{i=1}^p A_i \right) - I, \quad \Gamma_i = - \sum_{j=i+1}^p A_j. \quad (20.13)$$

Granger's representation theorem asserts that if the coefficient matrix  $\Pi$  has reduced rank  $\tau < k$ , then there exist  $k \times \tau$  matrices  $\alpha$  and  $\beta$  each with rank  $r$  such that  $\Pi = \alpha\beta'$  and  $\beta'y_t$  is stationary.  $r$  is the number of cointegrating relations (the *cointegrating rank*) and each column of  $\beta$  is the cointegrating vector. As explained below, the elements of  $\alpha$  are known as the *adjustment parameters* in the vector error correction model. Johansen's method is to estimate the  $\Pi$  matrix in an unrestricted form, then test whether we can reject the restrictions implied by the reduced rank of  $\Pi$ .

### The number of cointegrating relations

If you have  $k$  endogenous variables, each of which has one unit root, there can be from zero to  $k-1$  linearly independent, cointegrating relations. If there are no cointegrating relations, standard time series analyses such as the (unrestricted) VAR may be applied to the first-differences of the data. Since there are  $k$  separate integrated elements driving the series, levels of the series do not appear in the VAR in this case.

Conversely, if there is one cointegrating equation in the system, then a single linear combination of the levels of the endogenous series,  $\beta'y_{t-1}$ , should be added to each equation in the VAR. When multiplied by a coefficient for an equation, the resulting term,  $\alpha\beta'y_{t-1}$ , is referred to as an error correction term (see the discussion below). If there are additional cointegrating equations, each will contribute an additional error correction term involving a different linear combination of the levels of the series.

If there are exactly  $k$  cointegrating relations, none of the series has a unit root, and the VAR may be specified in terms of the levels of all of the series. Note that in some cases, the individual unit root tests will show that some of the series are integrated, but the Johansen tests show that the cointegrating rank is  $k$ . This contradiction may be the result of specification error.

### The Cointegrating Relations (Vector)

Each column of the  $\beta$  matrix gives an estimate of a cointegrating vector. The cointegrating vector is not identified unless we impose some arbitrary normalization. EViews adopts the normalization so that the  $r$  cointegrating relations are solved for the first  $r$  variables in the  $y_t$  vector as a function of the remaining  $k - r$  variables.

Note that one consequence of this normalization is that the normalized vectors which EViews provides will not, in general, be orthogonal, despite the orthogonality of the unnormalized coefficients.

### Deterministic Trend Assumptions

Your series may have nonzero means and deterministic trends as well as stochastic trends. Similarly, the cointegrating equations may have intercepts and deterministic trends. The asymptotic distribution of the LR test statistic for the reduced rank test does not have the usual  $\chi^2$  distribution and depends on the assumptions made with respect to deterministic trends. EViews provides tests for the following five possibilities considered by Johansen (see Johansen, 1995, pp. 80-84 for details):

1. Series  $y$  have no deterministic trends and the cointegrating equations do not have intercepts:

$$H_2(r): \Pi y_{t-1} + Bx_t = \alpha\beta'y_{t-1} \quad (20.14)$$

2. Series  $y$  have no deterministic trends and the cointegrating equations have intercepts:

$$H_1^*(r): \Pi y_{t-1} + Bx_t = \alpha(\beta'y_{t-1} + \rho_0) \quad (20.15)$$

3. Series  $y$  have linear trends but the cointegrating equations have only intercepts:

$$H_1(r): \Pi y_{t-1} + Bx_t = \alpha(\beta'y_{t-1} + \rho_0) + \alpha_\perp \gamma_0 \quad (20.16)$$

4. Both series  $y$  and the cointegrating equations have linear trends:

$$H^*(r): \Pi y_{t-1} + Bx_t = \alpha(\beta'y_{t-1} + \rho_0 + \rho_1 t) + \alpha_\perp \gamma_0 \quad (20.17)$$

5. Series  $y$  have quadratic trends and the cointegrating equations have linear trends:

$$H(r): \Pi y_{t-1} + Bx_t = \alpha(\beta'y_{t-1} + \rho_0 + \rho_1 t) + \alpha_\perp(\gamma_0 + \gamma_1 t) \quad (20.18)$$

where  $\alpha_\perp$  is the (non-unique)  $k \times (k - r)$  matrix such that  $\alpha'\alpha_\perp = 0$  and  $\text{rank}([\alpha, \alpha_\perp]) = k$ .

These five cases are nested from the most restrictive to the least restrictive, given any particular cointegrating rank  $r$

$$H_2(r) \subset H_1^*(r) \subset H_1(r) \subset H^*(r) \subset H(r) : \quad (20.19)$$

For each case, EViews tabulates the critical values for the reduced rank test as given by Osterwald-Lenum (1992), not those tabulated in Johansen and Juselius (1990). Note that the critical

values are available for up to 10 series and may not be appropriate for models that contain other deterministic regressors. For example, a shift dummy variable in the VAR implies a broken linear trend in the  $y$  series.

If you include 0–1 seasonal dummy variables in the VAR, this will affect both the mean and the trend of the  $y$  series. Johansen (1995, page 84) suggests using *centered* (orthogonalized) seasonal dummy variables, which only shift the mean without contributing to the trend. Centered seasonal dummy variables for quarterly and monthly series can be generated by the commands

```
series d_q=@seas(q)-1/4
series d_m=@seas(m)-1/12
```

for quarter  $q$  and month  $m$ , respectively.

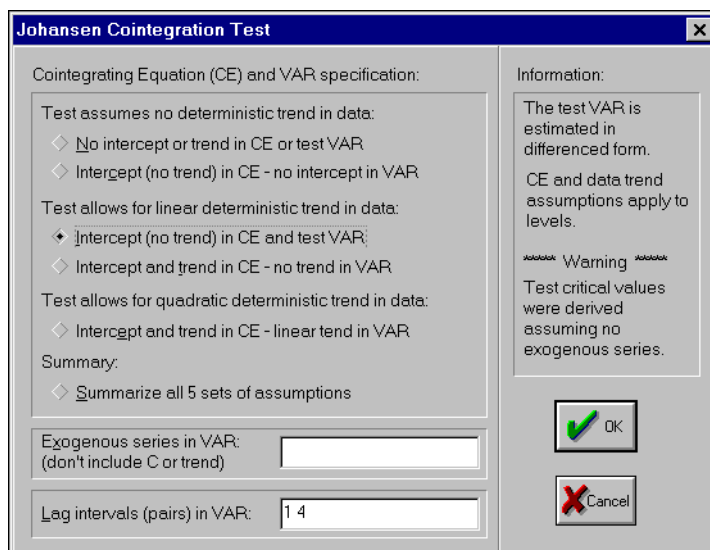
There are two dimensions you can vary while performing tests within this framework. You can assume one of the five cases listed above, and carry out tests for the cointegrating rank. Alternatively, you can fix the rank, and test which of the five cases describes the data best. These tests are standard  $\chi^2$  tests and are described in the “Technical Discussion” on page 502. EViews provides you with the option of summarizing all five cases, so you can examine all possible combinations of rank and intercept-trend.

Among the five intercept-trend cases, EViews uses Equation (20.16) as the default (case 3: series has linear trend but the cointegrating relations have only an intercept). This choice is based on the assumption that long-run equilibrium conditions (such as the relation between income and consumption) do not have trends. In choosing your final model of the data, you should be guided by economic interpretations as well as the statistical criteria. You should look at the normalized cointegrating equations to see if they correspond to your beliefs about long-run relations among the variables. Sometimes it is helpful to rearrange the order of the variables; a reordering does not affect the substantive results.

## How to Perform a Johansen Test

To carry out the Johansen test, select **View/Cointegration Test...** on the group or VAR toolbar. Note that since this is a test for cointegration, this test is only valid when you are working with series that are known to be nonstationary. You may wish first to apply unit root tests to each series in the VAR (see Chapter 13, page 298 for details on carrying out unit root tests in EViews).

The Cointegration Test dialog asks you to provide information about the test:



## Specification

The first step is to choose one of the options specifying the type of deterministic trends that are present in the data. The first five options provide particular alternatives for whether an intercept or trend term should be included in the specification of the cointegrating equations.

The sixth option runs the Johansen procedure under all five sets of assumptions. Since this selection estimates a large number of specifications, only a summary table of results will be displayed. You must select one of the other options to see the full output of the test.

## Exogenous Variables

The dialog also allows you to specify exogenous variables, such as seasonal dummies, to include in the test VARs. The constant and linear trend should *not* be listed in this dialog; these should be chosen from the specification options. If you do include exogenous variables, you should be aware that the critical values reported by EViews do not account for their inclusion.

## Lag Intervals

You should specify the lags of the test VAR as pairs of intervals. *Note that in contrast to some other statistical packages, the lags are specified as lags of the first differenced terms, not in terms of the levels.* For example, if you type 1 4 in the field, the test VAR regresses  $\Delta y_t$  on  $\Delta y_{t-1}$ ,  $\Delta y_{t-2}$ ,  $\Delta y_{t-3}$ ,  $\Delta y_{t-4}$  and other exogenous variables that you have specified. Some other programs specify the lags in levels; if you type 1 4 in EViews, the highest lag in level is 5. To run a cointegration test with one lag in the *levels*, type “0 0” in the edit field.

## Interpreting the Results of a Johansen Test

As an example, the output for a four-variable system used in Johansen and Juselius (1990) for the Danish data is shown below. The four variables are

LRM: log real money  
 LRY: log real income  
 IBO: long term interest rate (bond rate)  
 IDE: short term interest rate (deposit rate)

The test assumes no trend in the series with an intercept in the cointegration relation (second specification in the dialog), includes three orthogonalized seasonal dummy variables, and uses two lags in *levels* which is specified as “1 1” in the edit field box.

### The number of cointegrating relations

The first part of the table performs the trace test for the number of cointegrating relations:

Date: 10/16/97 Time: 22:33  
 Sample: 1974:1 1987:3  
 Included observations: 53  
 Test assumption: No deterministic trend in the data  
 Series: LRM LRY IBO IDE  
 Exogenous series: D1 D2 D3  
 Warning: Critical values were derived assuming no exogenous series  
 Lags interval: 1 to 1

Eigenvalue	Likelihood Ratio	5 Percent Critical Value	1 Percent Critical Value	Hypothesized No. of CE(s)
0.433165	49.14436	53.12	60.16	None
0.177584	19.05691	34.91	41.07	At most 1
0.112791	8.694964	19.96	24.60	At most 2
0.043411	2.352233	9.24	12.97	At most 3

\*(\*\*) denotes rejection of the hypothesis at 5%(1%) significance level

L.R. rejects any cointegration at 5% significance level

The eigenvalues are presented in the first column, while the second column (Likelihood Ratio) gives the LR test statistic:

$$Q_{\tau} = -T \sum_{i=r+1}^k \log(1 - \lambda_i) \quad (20.20)$$

for  $\tau = 0, 1, \dots, k-1$  where  $\lambda_i$  is the  $i$ -th largest eigenvalue.  $Q_{\tau}$  is termed the *trace statistic* and is used to test  $H_1(\tau)$  against  $H_1(k)$ .

To determine the number of cointegrating relations  $r$ , subject to the assumptions made about the trends in the series, we can proceed sequentially from  $r = 0$  to  $r = k-1$  until we fail to

reject. The first row in the upper table tests the hypothesis of no cointegration, the second row tests the hypothesis of one cointegrating relation, the third row tests the hypothesis of two cointegrating relations, and so on, all against the alternative hypothesis of full rank, *i.e.* all series in the VAR are stationary.

The trace statistic does not reject any of the hypotheses at the 5% level. Note that EViews displays the critical values for the trace statistic reported by Osterwald-Lenum (1992), not those tabulated in Johansen and Juselius (1990). Johansen also proposes an alternative LR test statistic, known as the *maximum eigenvalue statistic*, which tests  $H_1(\tau)$  against  $H_1(\tau + 1)$ . The maximum eigenvalue statistic can be computed from the trace statistic as

$$Q_{\max} = -T \log(1 - \lambda_{\tau+1}) = Q_{\tau} - Q_{\tau+1} \quad (20.21)$$

EViews does not provide critical values for the maximum eigenvalue statistic; critical values for this statistic are tabulated in Osterwald-Lenum (1992).

### The Cointegrating Equations

Below the results of the cointegration rank tests, EViews provides the estimates of the cointegrating vector or relations. Although EViews displays all possible  $k - 1$  cointegrating relations, you are most likely to be interested in the first  $r$  estimates, where  $r$  is determined by the LR tests.

The cointegrating vector is not identified unless we impose some arbitrary normalization. EViews adopts a normalization such that the first  $r$  series in the  $y_t$  vector are normalized to an identity matrix. The normalized cointegrating relation assuming one cointegrating relation  $\tau = 1$  is given by

Normalized Cointegrating Coefficients: 1 Cointegrating Equation(s)				
LRM	LRY	IBO	IDE	C
1.000000	-1.032949 (0.13366)	5.206919 (0.66493)	-4.215880 (1.17916)	-6.059932 (0.83100)
Log likelihood      669.1154				

which can be written as

$$LRM - 1.03LRY + 5.21IBO - 4.22IDE - 6.06 \quad (20.22)$$

The number in parentheses under the estimated coefficients are the asymptotic standard errors. Some of the normalized coefficients will be shown without standard errors. This will be the case for coefficients that are normalized to 1.0 and for coefficients that are not identified. In the latter case, the coefficient (usually 0) is the result of an arbitrary identifying assumption.

The appearance of the normalized cointegrating relations depends on how you order the series in the VAR. For example, if you want a relation with a unit coefficient on LRY, specify the LRY



series as the first series in your VAR. Of course, nothing of substance is affected since any linear combination of the cointegrating relations is also a cointegrating relation.

### Summary of Five Tests

If you are not sure about the deterministic trends in the data, you may choose the summary of the cointegration tests under all five models. The output displays the log likelihood and the two information criteria under each model as a model selection guide. See Appendix F for a discussion of the use of information criteria in model selection.

The results of the LR tests shown at the bottom of the table are at the 5% significance level.

## Estimating a VEC in EViews

An unrestricted VAR does not assume the presence of cointegration. If you wish to impose cointegrating restrictions among the variables in the VAR, you should use a vector error correction (VEC) model.

As the VEC specification only applies to cointegrated series, you should run the Johansen cointegration test prior to VEC specification as described above. This allows you to confirm that the variables are cointegrated and to determine the number of cointegrating equations.

To set up a VEC, click the **Estimate** button in the VAR toolbar and choose the **Vector Error Correction** specification.

You should provide the same information as for an unrestricted VAR and you may optionally specify exogenous variables. However, the specification of the exogenous intercepts and trends should be chosen from the five models discussed above. This choice should be the same as in the cointegration test.

*It is important to note that the lag specification that EViews prompts you to enter refers to lags of the first difference terms in the*

**Vector Error Correction**

VAR specification:

- ☐ Unrestricted VAR
- ☒ Vector Error Correction

Lag intervals (as range pairs):

1 12

Sample:

1959:01 1989:12

Series to include in VEC:

Endogenous:

FF TB3

Exogenous: (don't include C or trend)

Cointegrating equation (CE):

VAR assumes no deterministic trend:

- ☐ No intercept or trend in CE
- ☐ Intercept (no trend) in CE

VAR assumes linear trend in data:

- ☒ Intercept (no trend) in CE
- ☐ Intercept and trend in CE

VAR assumes quadratic trend in data:

- ☐ Intercept and trend in CE

Number of CE's:

1

OK Cancel

VEC. For example, 1 1 specifies a model involving a regression of the first differences on one lag of the *first difference*.

You must also specify the number of cointegrating equations in the VEC model. This number should be determined from the cointegration test. The maximum number of cointegrating equations is one less than the number of endogenous variables in the VAR.

To estimate the VEC, click **OK**. Estimation of a VEC model proceeds by first determining one or more cointegrating equations using the Johansen procedure. The first difference of each endogenous variable is then regressed on a one period lag of the cointegrating equation(s) and lagged first differences of all of the endogenous variables in the system.

## Working with a VEC

Working with a VEC is analogous to working with a VAR, so we refer you to the discussion for the VAR above. Note that standard errors for the impulse response functions are not available for the VEC.

### Obtaining the Coefficients of a VEC

As explained in the VAR section, estimated coefficients of VARs can be accessed by reference to a two dimensional array C. Similarly, estimated coefficients of the cointegrating vector and the adjustment parameters can be accessed by reference to the two dimensional arrays B and A, respectively.

To see the correspondence between each element of A, B, C and the estimated coefficients, select **View/Representations** from the VAR toolbar. The top part of the view is presented below:

```
Estimation Command:
=====
EC 1 4 IP M1 PPI TB3 @ C

VAR Model:
=====
D(IP) = A(1,1)*(B(1,1)*IP(-1) + B(1,2)*M1(-1) + B(1,3)*PPI(-1) + B(1,4)*TB3(-1) + B(1,5)) +
A(1,2)*(B(2,1)*IP(-1) + B(2,2)*M1(-1) + B(2,3)*PPI(-1) + B(2,4)*TB3(-1) + B(2,5)) + C(1,1)*D(IP(-1)) +
C(1,2)*D(IP(-2)) + C(1,3)*D(IP(-3)) + C(1,4)*D(IP(-4)) + C(1,5)*D(M1(-1)) + C(1,6)*D(M1(-2)) +
C(1,7)*D(M1(-3)) + C(1,8)*D(M1(-4)) + C(1,9)*D(PPI(-1)) + C(1,10)*D(PPI(-2)) + C(1,11)*D(PPI(-3)) +
C(1,12)*D(PPI(-4)) + C(1,13)*D(TB3(-1)) + C(1,14)*D(TB3(-2)) + C(1,15)*D(TB3(-3)) + C(1,16)*D(TB3(-
4)) + C(1,17)

D(M1) = A(2,1)*(B(1,1)*IP(-1) + B(1,2)*M1(-1) + B(1,3)*PPI(-1) + B(1,4)*TB3(-1) + B(1,5)) +
A(2,2)*(B(2,1)*IP(-1) + B(2,2)*M1(-1) + B(2,3)*PPI(-1) + B(2,4)*TB3(-1) + B(2,5)) + C(2,1)*D(IP(-1)) +
C(2,2)*D(IP(-2)) + C(2,3)*D(IP(-3)) + C(2,4)*D(IP(-4)) + C(2,5)*D(M1(-1)) + C(2,6)*D(M1(-2)) +
C(2,7)*D(M1(-3)) + C(2,8)*D(M1(-4)) + C(2,9)*D(PPI(-1)) + C(2,10)*D(PPI(-2)) + C(2,11)*D(PPI(-3)) +
C(2,12)*D(PPI(-4)) + C(2,13)*D(TB3(-1)) + C(2,14)*D(TB3(-2)) + C(2,15)*D(TB3(-3)) + C(2,16)*D(TB3(-
4)) + C(2,17)
```

The first index to A is the equation number of the VEC, while the second index to A is the number of the cointegrating equation. For example, A(2,1) is the adjustment coefficient of the first cointegrating equation in the second equation of the VEC.

The first index to B is the number of the cointegrating equation, while the second index to B is the variable number in the cointegrating equation. For example, B(2,1) is the coefficient of the first variable in the second cointegrating equation.

The first index to C is the equation number of the VEC, while the second index to C is the variable number of the first-differenced regressor of the VEC. For example, C(2,1) is the coefficient of the first first-differenced regressor in the second equation of the VEC.

You can access each element of the coefficient by referring to the name of the VEC followed by a dot and coefficient element:

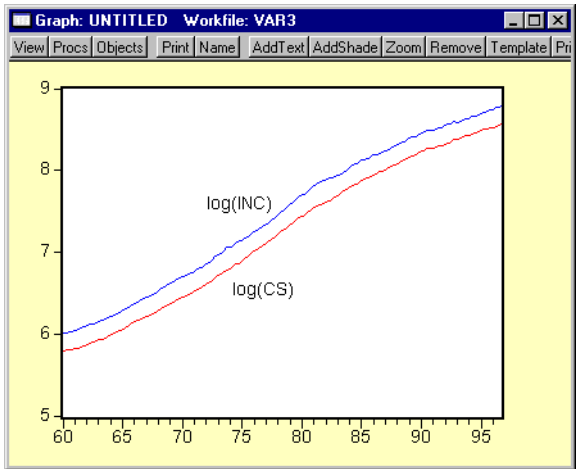
```
var01.a(2,1)
var01.b(2,1)
var01.c(2,1)
```

## Forecasting from a VAR or VEC

To calculate a forecast from a VAR or Vector Error Correction model, click on **Procs/Make Model** on the VAR toolbar. You will see a model window with the VAR or VEC in it. You can make any changes you want, including modifying the ASSIGN statement, and then push the **Solve** button on the model's toolbar to calculate a forecast. See Chapter 23 for further discussion on forecasting using models.

## Illustration

As an illustration, we estimate a bivariate VAR of the log of personal consumption expenditure (CS) and the log of personal income (INC) using quarterly data over the period 60.1–96.4. The two series are plotted below:



The two series are clearly drifting together, suggesting cointegration. To establish cointegration, we must first check whether each series is integrated and contains a unit root. The results of the augmented Dickey-Fuller test with four lags and a constant and linear trend in the test equation are presented below:

	$\log(\text{CS})$	$\Delta \log(\text{CS})$	$\log(\text{INC})$	$\Delta \log(\text{INC})$
ADF $t$ -stat	-0.52	-3.67	-0.92	-3.78

The 5% critical value is -3.44 and the tests indicate that both series contain a unit root.

### Choosing the Lag Order of a VAR

To estimate a VAR of  $\log(\text{CS})$  and  $\log(\text{INC})$ , click **Quick/Estimate VAR...** or type `var` in the command window. To fit a fourth order VAR, fill in the dialog as follows:

Click **OK** to estimate the VAR. After estimation, you can select **View/Residual Graphs** to plot the residuals of each equation in the VAR.

The lag order of the VAR is often selected somewhat arbitrarily, with standard recommendations suggesting that you set it long enough to ensure that the residuals are white noise. How-

ever, if you choose the lag length too large, the estimates become imprecise. You can use a likelihood ratio (LR) test to test the appropriate lag length. To carry out the LR test, estimate the VAR twice, each with different lags. The estimation results for lag orders 4 (top) and 2 (bottom) are displayed below:

R-squared	0.999946	0.999933
Adj. R-squared	0.999943	0.999929
Sum sq. resids	0.006336	0.007886
S.E. equation	0.006751	0.007532
Log likelihood	534.3456	518.1447
Akaike AIC	-9.937142	-9.718211
Schwarz SC	-9.754879	-9.535948
Mean dependent	7.216626	7.452695
S.D. dependent	0.897423	0.894048
Determinant Residual Covariance		1.62E-09
Log Likelihood		1077.958
Akaike Information Criteria		-14.32376
Schwarz Criteria		-13.95923

R-squared	0.999942	0.999923
Adj. R-squared	0.999940	0.999921
Sum sq. resids	0.006853	0.009006
S.E. equation	0.006923	0.007936
Log likelihood	528.5356	508.3217
Akaike AIC	-9.912682	-9.639522
Schwarz SC	-9.811425	-9.538265
Mean dependent	7.216626	7.452695
S.D. dependent	0.897423	0.894048
Determinant Residual Covariance		1.90E-09
Log Likelihood		1065.907
Akaike Information Criteria		-14.26901
Schwarz Criteria		-14.06649

The LR test statistic for the hypothesis of lag 2 against lag 4 can be computed as:

$$LR = -2(l_2 - l_4) = -2(1065.907 - 1077.958) = 24.102 \quad (20.23)$$

where  $l$  is the log likelihood reported at the very bottom of the table. The LR test statistic is asymptotically distributed  $\chi^2$  with degrees of freedom equal to the number of restrictions under test. In this case, there are 8 zero restrictions to move from VAR(4) to VAR(2) so the LR test statistic is asymptotically distributed  $\chi^2(8)$ . To carry out the test, you can compare the test statistic with the critical values from the  $\chi^2$  table. Alternatively, you can compute the  $p$ -value in EViews by entering the commands:

```
scalar pval = 1-@cchisq(24.102,8)
show pval
```

The result 0.002 is displayed on the status line at the bottom of the EViews window and we reject the null hypothesis of 2 lags. The Akaike information criterion also selects VAR(4) in favor of VAR(2) but the Schwarz information criterion selects the parsimonious VAR(2).

Note: to compute the LR test appropriately, you must estimate the two VARs using the same sample period. The VAR with the longer lag will have a shorter sample, so you should set the sample period to the sample estimated by the longer VAR.

## Testing for Cointegration

Given that each series has a unit root, we now test whether CS and INC are cointegrated over the sample period and if so, what the cointegrating relation is. The estimated cointegrating relation will give us an estimate of the long run propensity to consume.

To test cointegration, click **View/Cointegration Test...** and fill in the dialog. We use the VAR with lag 4 and assume that the series contain linear trends but the cointegrating relation only includes a constant (the default third model). The results are:

Date: 10/16/97 Time: 23:19  
Sample: 1960:1 1996:4  
Included observations: 147  
Test assumption: Linear deterministic trend in the data  
Series: LOG(CS) LOG(INC)  
Lags interval: 1 to 4

Eigenvalue	Likelihood Ratio	5 Percent Critical Value	1 Percent Critical Value	Hypothesized No. of CE(s)
0.145518	25.20440	15.41	20.04	None **
0.014098	2.087119	3.76	6.65	At most 1

\*(\*\*) denotes rejection of the hypothesis at 5%(1%) significance level

L.R. test indicates 1 cointegrating equation(s) at 5% significance level

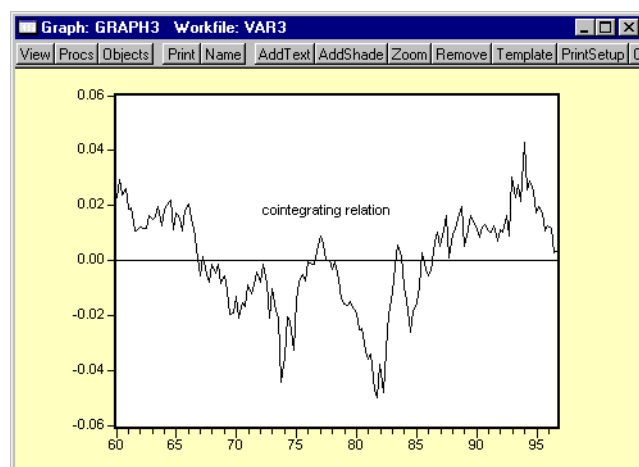
The LR test rejects the hypothesis of no cointegration but not the hypothesis of at most one cointegration relation. The estimated cointegrating relation is:

Normalized Cointegrating Coefficients: 1 Cointegrating Equation(s)

LOG(CS)	LOG(INC)	C
1.000000	-1.002850 (0.00320)	0.257338
Log likelihood	1076.042	

As expected, the estimated long run marginal propensity to consume is not significantly different from one. You can plot the cointegrating relation by the command

```
plot log(cs) -1.002850*log(inc) + 0.257338
```



You can draw the zero line by double clicking in the white area of the graph and choosing **Graph Attributes: Zero line**. The graph shows that consumption was less than the long run level during most of the 70's.

## Technical Discussion

Consider a VAR of order  $p$

$$y_t = Bx_t + A_1y_{t-1} + \dots + A_p y_{t-p} + \epsilon_t \quad (20.24)$$

where  $y_t$  is a  $k$  vector of endogenous variables,  $x_t$  is a  $d$  vector of exogenous variables, and  $\epsilon_t$  is a vector of innovations with variance matrix  $\Omega$ .

Assuming the VAR is invertible, we obtain the MA representation as

$$\begin{aligned} \tilde{y}_t &= A_1 \tilde{y}_{t-1} + \dots + A_p \tilde{y}_{t-p} + \epsilon_t \\ &= (I - A_1 L - \dots - A_p L^p)^{-1} \epsilon_t \\ &= (I + \Psi_1 L + \Psi_2 L^2 + \dots) \epsilon_t \end{aligned} \quad (20.25)$$

where  $\tilde{y}$  is the residual from regressing  $y$  on the exogenous variables  $x$ . The coefficients  $\Psi$  of the MA representation are obtained as follows. The VAR coefficients  $A$  and the MA coefficients  $\Psi$  must satisfy the relation

$$\begin{aligned} (I - A_1 L - \dots - A_p L^p)(I + \Psi_1 L + \Psi_2 L^2 + \dots) &= I \\ I + C_1 L + C_2 L^2 + \dots &= I \end{aligned} \quad (20.26)$$

where  $C_1 = C_2 = \dots = 0$ . These conditions on  $C_s$  recursively define the MA coefficients as

$$\begin{aligned} \Psi_1 &= A_1 \\ \Psi_2 &= A_1 \Psi_1 + A_2 \\ &\vdots \\ \Psi_s &= A_1 \Psi_{s-1} + A_2 \Psi_{s-2} + \dots + A_p \Psi_{s-p} \end{aligned} \quad (20.27)$$

## Impulse Response

The orthogonalized impulse response at lag  $s$  is given by  $\Psi_s P$  where  $P$  is a  $k \times k$  lower triangular matrix with the standard deviations of  $\epsilon$  along the main diagonal such that  $PP' = \Omega$ . The row  $i$  column  $j$  element of  $\Psi_s P$  is the effect of a one standard deviation *orthogonalized* shock to  $y_{jt}$  on  $y_{i(t+s)}$ , holding all other shocks at all dates constant.

The analytic asymptotic standard errors of the impulse response functions are computed according to the formula given in Hamilton (1994a, page 339).



The Monte Carlo standard errors are computed as follows. At each repetition, we draw a random sample from the asymptotic distribution of the VAR coefficients. The asymptotic distribution of the VAR coefficients is given in Hamilton (1994a, Proposition 11.2.). From these simulated coefficients we compute the impulse response functions. After repeating this process  $m$  times, we construct the 95 % confidence interval by the percentile method. The standard errors reported in the table are the standard deviations of the simulated impulse responses across  $m$  replications.

## Variance Decomposition

The  $s$ -period ahead forecast error from a VAR is

$$\epsilon_{t+s} + \Psi_1 \epsilon_{t+s-1} + \Psi_2 \epsilon_{t+s-2} + \dots + \Psi_{s-1} \epsilon_{t+1} \quad (20.28)$$

with mean squared error

$$\begin{aligned} & \Omega + \Psi_1 \Omega \Psi_1' + \dots + \Psi_{s-1} \Omega \Psi_{s-1}' \\ &= (PP' + \Psi_1 PP' \Psi_1' + \dots + \Psi_{s-1} PP' \Psi_{s-1}') \\ &= \sum_{j=1}^k (p_j p_j' + \Psi_1 p_j p_j' \Psi_1' + \dots + \Psi_{s-1} p_j p_j' \Psi_{s-1}') \end{aligned} \quad (20.29)$$

where  $p_j$  is the  $j$ -th column of  $P$ . The expression in parentheses is the contribution of the  $j$ -th orthogonalized innovation to the mean squared error of the  $s$ -period ahead forecast.

## Nested Tests of Deterministic Trends under Cointegration

For a fixed number of cointegrating relations  $r$ , the five models are nested in the following order:

$$H_2(r) \subset H_1^*(r) \subset H_1(r) \subset H^*(r) \subset H(r) \quad (20.30)$$

The  $\chi^2$  test statistics have the form

$$\begin{aligned} T \sum_{i=r+1}^k \log((1 - \lambda_i)/(1 - \lambda_i^*)) &\sim \chi^2(k-r) \\ T \sum_{i=1}^r \log((1 - \lambda_i)/(1 - \lambda_i^*)) &\sim \chi^2(r) \end{aligned} \quad (20.31)$$

where  $\lambda_i$  and  $\lambda_i^*$  are the  $i$ -th largest eigenvalues under the models  $H$  and  $H^*$ , respectively. The test for  $H_1^*(r)$  against  $H_2(r)$  and the test for  $H^*(r)$  against  $H_1(r)$  use the  $\chi^2(r)$  test. The remaining two tests use the  $\chi^2(k-r)$  form.

## Standard Errors of the Cointegrating Equations

The formula for the asymptotic standard errors can be found in Theorem 5.1, Equation (5.6) in Johansen (1991) with  $H$  set to the identity matrix (unrestricted) or Theorem 13.3, Equation (13.12) in Johansen (1995). The standard errors are reported only for the variables that are identified under the chosen normalization.

## Commands

To declare a VAR, follow the `var` command with a name for the VAR:

```
var finvar
```

If you only type `var` without a name, EViews brings up the estimation dialog of an untitled VAR.

You can also define and estimate a VAR in one command. For example,

```
var finvar.ls 1 4 gdp m1 gov
```

declares a VAR named FINVAR and estimates a trivariate unrestricted VAR with 4 lags and a constant. See the *Command and Programming Reference* for a complete list of commands and options for VAR objects.

## Chapter 21. State Space Models and the Kalman Filter

---

Dynamic systems can be represented in a general form known as the state space form. Many time series models, including the classical linear regression model and ARIMA models, can be written and estimated as special cases of a state space specification.

There are two main benefits to representing a dynamic system in state space form. First, the state space allows unobserved variables (known as the *state* variables) to be incorporated into, and estimated along with, the observable model. Second, state space models can be estimated using a powerful recursive algorithm known as the *Kalman filter*. The Kalman filter is used both to evaluate the likelihood function and to forecast and smooth the unobserved state variables.

State space models have been applied in the econometrics literature to model unobserved variables: (rational) expectations, measurement errors, missing observations, permanent income, unobserved components (cycles and trends), and the natural rate of unemployment. The Kalman filter algorithm has been used to compute exact, finite sample forecasts for Gaussian ARMA models, multivariate (vector) ARMA models, MIMIC (multiple indicators and multiple causes), Markov switching models, and time varying (random) coefficient models. Extensive surveys of applications of state space models in econometrics can be found in Hamilton (1994a, Chapter 13; 1994b) and Harvey (1989, Chapters 3, 4).

EViews estimates both single and multiple equation dynamic models using a state space object. The state space object provides you with tools for specifying, estimating, and working with the results of your dynamic system.

### State Space Models

We present here a brief discussion of the specification and estimation of a state space model. Those desiring greater detail are directed to Hamilton (1994a, Chapter 13) and Harvey (1993, Chapter 4).

#### The State Space Representation

The state space representation of the dynamics of an  $n \times 1$  vector  $y$  is given by the following system of equations:

$$\begin{aligned}y_t &= A'x_t + H(z_t)'\xi_t + w_t \\ \xi_{t+1} &= F\xi_t + v_{t+1}\end{aligned}\tag{21.1}$$

for where  $A'$ ,  $H'$ , and  $F$  are matrices of dimension  $n \times k$ ,  $n \times r$ , and  $r \times r$ , respectively.  $x$  is a  $k \times 1$  vector of exogenous or predetermined variables and  $\xi$  is a  $r \times 1$  vector of possibly unobserved state variables. The first equation is known as the observation (or measurement) equation and the second is known as the state (or transition) equation. The disturbance vectors,  $w_t$  and  $v_t$  are assumed to be independent white noise with

$$\text{var}(w_t) = R, \quad \text{var}(v_t) = Q, \quad E(w_s v_t) = 0 \quad (21.2)$$

for all  $s$  and  $t$ ,  $s \neq t$

### Estimation by the Kalman filter

Given observations  $(y_t, x_t)$  for  $t = 1, 2, \dots, T$ , our goal is to estimate the parameters  $A$ ,  $H$ ,  $F$ ,  $R$ ,  $Q$ , and make inferences about the state vector  $\xi$ . The Kalman filter is a recursive algorithm for sequentially updating the state vector given past information. More technically, it is an algorithm for calculating linear least squares forecasts of the state vector given data observed up to date  $t$ . The state vector  $\xi$  and its mean squared error  $P_t = E[(\xi_t - \hat{\xi}_t)(\xi_t - \hat{\xi}_t)']$  are recursively estimated by

$$\begin{aligned} \xi_{t|t} &= F\xi_{t-1|t-1} + S_{t-1}H(H'S_{t-1}H + R)^{-1}(y_t - A'x_t - H'F\xi_{t-1|t-1}) \\ P_{t|t} &= S_{t-1} - S_{t-1}H(H'S_{t-1}H + R)^{-1}H'S_{t-1} \end{aligned} \quad (21.3)$$

where  $S_{t-1} = FP_{t-1|t-1}F' + Q$ , and  $\xi_{s|v}$  is the forecast of the state vector at time period  $s$ , given information available at time  $v$ . Note that the recursion for  $P$  does not depend on the filtered state vector  $\xi_{t-1|t-1}$ , or on the observed data  $(y_t, x_t)$ .

To implement the Kalman filter, we must specify the starting values and replace the unknown matrices by their estimates. By default, EViews obtains starting values by treating these matrices as fixed coefficients and estimating them using OLS.

Once we obtain starting values, the parameters  $A$ ,  $H$ ,  $F$ ,  $R$ ,  $Q$  are estimated by maximizing the log likelihood function under the assumption that the distribution of  $y_t$ , conditional on  $x_t$  and past values of  $(y_t, x_t)$ , is multivariate normal (Gaussian).

See “Technical Discussion” on page 522 below for a more detailed description of the iteration algorithm.

## Specifying State Space Models

The first step in specifying and estimating a state space model is to create a state space object. Select **Objects/New Object/StateSpace** from the main toolbar or type `sspace` in the command window. An empty state space window will open.

The next step is to specify the model. EViews estimates the following broad classes of state space models:

- Time-varying parameter models.
- Recursive Coefficient Models.
- ARMA models.

EViews has special “auto-specification” features to guide you in creating standard forms of these models. Simply press the **AutoSpec** button on the state space toolbar and select **Time-Varying Parameter...**, **Recursive Estimation...**, or **ARMA Estimation...** as appropriate. Specialized dialogs will open to guide you through the specification process.

If you wish, you can customize your auto-specified model, or simply create one of these specifications from scratch.

### Time-Varying Parameter Models

The class of models that can be estimated with this selection have the following form:

$$\begin{aligned} y_t &= z_t' \alpha + x_t' \beta_t + w_t \\ \beta_{t+1} - \bar{\beta} &= F(\beta_t - \bar{\beta}) + v_{t+1} \end{aligned} \quad (21.4)$$

$z$  is the set of regressors with fixed coefficients, while  $x$  is the set of regressors with random coefficients. The vector of random coefficients  $\xi_t = \beta_t - \bar{\beta}$  is the state vector, where  $\bar{\beta}$  is the steady-state value of the coefficient vector if the process is stationary. The state equation describes the dynamics of the coefficients  $\beta$ , which are assumed to follow a vector-AR(1) process.

To specify a time-varying parameter model, press **AutoSpec** on the state space toolbar, choose **Time-Varying Parameter...** and fill in the specification dialog box.

First, you should fill in the edit boxes with the names of the dependent variables for the observation equation or equations, the names of any regressors for which you would like to have fixed coefficients, and the names of any regressors for which you would like to have random coefficients (state variables). You can enter multiple dependent variables, regressors and random coefficients.

Next, select a structure for your state equation using the **Random coefficient assumption** radio buttons.

- **Random walk:** use this option if you believe that shocks to the random coefficients persist indefinitely.
- **AR(1) and constant mean:** use this option if you believe that shocks to the random coefficients have some persistence, but that the coefficients eventually return to their mean values.
- **Constant mean:** use this option if you believe that shocks to the random coefficients have no persistence from period to period.

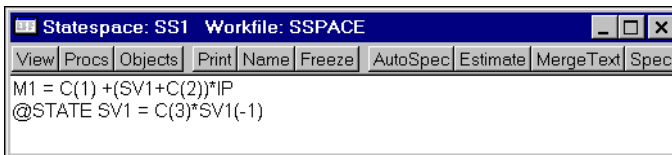
Lastly, you should enter a name in the edit box for **Base Name for coef series**. EViews will use this base name when creating the estimated state series. For example, if you enter SER in the edit box, EViews will use the next unused name of the form SER1, SER2, etc.

For example, suppose you want to estimate the following single equation model:

$$\begin{aligned} M1_t &= \alpha + IP_t \beta_t + w_t \\ \beta_{t+1} - \bar{\beta} &= \phi(\beta_t - \bar{\beta}) + v_{t+1} \end{aligned} \quad (21.5)$$

where the constant is a fixed parameter and the slope coefficient follows an AR(1) process.

Press **AutoSpec** and select **Time-varying parameter...** Enter M1 in the dependent variable field, C in the constant coefficient edit box, and IP in the state coefficient box. Since the state equation is an unrestricted AR(1), we select **AR(1) and constant mean**. In the example above, we have entered SV in the base name edit field. Click **OK** and EViews creates and displays the specification in the object window.



Below, we describe in greater detail the characteristics of a state space specification. For now, however, note that the first equation in this window is the observation equation. C(1) indicates that the fixed

intercept  $\alpha$  will be stored in the first row of the coefficient vector C. Since the state vector is defined as  $\xi_t = \beta_t - \bar{\beta}$ , the random slope coefficient can be written as  $\beta_t = \bar{\beta} + \xi_t$ . The state vector,  $\xi$ , will be stored as a series named SV1 and the mean,  $\bar{\beta}$ , will be stored in the second row of the coefficient vector C. The @state keyword indicates that the second equation is the state equation. The estimated AR(1) coefficient of the state equation will be stored in the third row of the coefficient vector C.

## Recursive Estimation

The class of models that can be estimated with this option have the following form:

$$\begin{aligned}y_t &= z_t' \beta_t + w_t \\ \beta_{t+1} &= \beta_t\end{aligned}\tag{21.6}$$

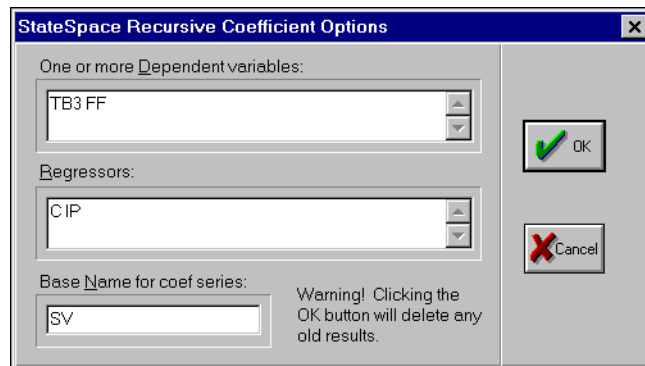
where the state vector is the recursive coefficient vector  $\beta_t$ . Recursive estimation restricts the state equation to have zero variance and unit coefficient. While the parameters in time-varying estimation are stochastic, the parameters in recursive estimation are deterministic.

Note that single equation recursive estimation can be performed using the state space object, or using the Recursive Least Squares view of an equation object. If you wish to recursively estimate a system of equations, you will need to specify and estimate a state space model.

To specify a recursive estimation model, press **AutoSpec**, select **Recursive Estimation...** and fill in the specification dialog box. For example, suppose you want to recursively estimate the following two equation model:

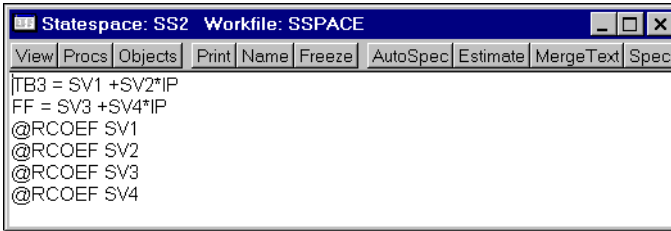
$$\begin{aligned}TB3_t &= \alpha_{1t} + \alpha_{2t}IP_t + w_{1t} \\ FF_t &= \beta_{1t} + \beta_{2t}IP_t + w_{2t}\end{aligned}\tag{21.7}$$

To specify this model in EViews, press **AutoSpec** and select **Recursive Estimation....** Note that EViews opens a dialog that is designed specifically for recursive estimation.



You will need to enter the dependent variable or variables for the observable equation(s), as well as the regressors. You must also specify a base name to be used for the recursive coefficients.

When you have completed the dialog, the state space window displays the specification:



The first two equations are the observation equations. The recursive intercept for the first equation is saved as a series named SV1, the recursive coefficient in the first equation is saved as SV2, and so on. The last four equations identify the state equations.

## ARMA Estimation

The class of models that can be estimated with this option have the following form:

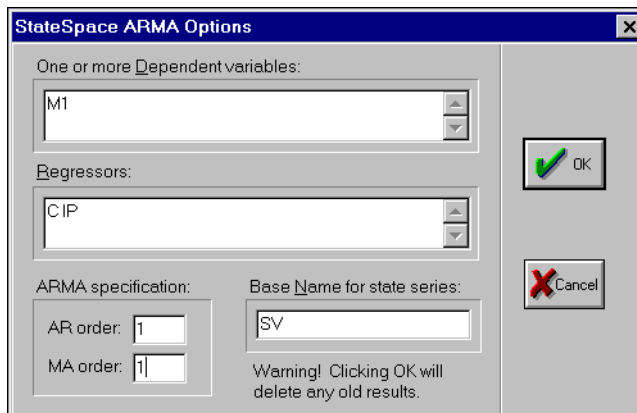
$$\begin{aligned} y_t &= A'x_t + H'\xi_t \\ \xi_{t+1} &= F\xi_t + v_{t+1} \end{aligned} \quad (21.8)$$

ARMAX models (ARMA models with exogenous variables) are written in state space form by restricting the observation equation to have no error term. The state vector is part of the disturbance term and the shocks  $v$  in the state equation are the innovations. The coefficient matrix  $H$  in the observation equation are the MA parameters, while the coefficient matrix  $F$  in the state equation are the AR parameters.

To specify an ARMA state space model, select **AutoSpec/ARMA Estimation** and fill in the specification dialog box. For example, suppose you want to estimate the following ARMA(1,1) model:

$$\begin{aligned} M1_t &= \beta_1 + \beta_2 IP_t + u_t \\ u_t &= \phi u_{t-1} + \epsilon_t + \theta \epsilon_{t-1} \end{aligned} \quad (21.9)$$

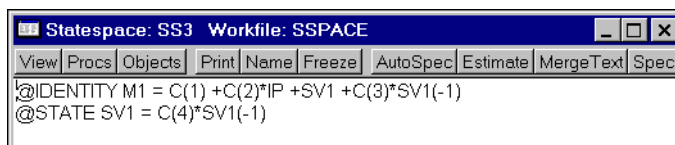
To specify this model in EViews, fill in the StateSpace ARMA dialog as follows:





Simply list the name of the dependent variable(s) and exogenous regressors in the dialog box and choose the order of the AR and MA terms. **Base Name for coef series** option allows you to choose a name for the recursive coefficient series to store in the workfile.

After you have completed specifying the model, click **OK** and the state space window displays the specification:



The first equation is the observation equation. The `@identity` keyword indicates that there are no error terms in the observation equation. The second equation, which begins with the `@state` keyword, is the state equation. To see that this is indeed an ARMA(1,1) model, eliminate the state variable from the two equations to get

$$M1_t - c(1) - c(2)IP_t = c(4)(M1_{t-1} - c(1) - c(2)IP_{t-1}) + \epsilon_t + c(3)\epsilon_{t-1} \quad (21.10)$$

This expression shows that  $C(3)$  is the MA(1) coefficient and  $C(4)$  is the AR(1) coefficient.

## General Specifications

You can specify your model directly in the state space window without using the **AutoSpec** menu. EViews will try to fit your specification into one of the three types of models in the **AutoSpec** menu. Therefore you must first determine whether your model fits into one of the three types. If it does, you should conform to the syntax specific to each type of model.

### Recursive models

#### Observation equations

The coefficients in the observation equations are the state variables. Specify the equations by formula, just as you would in a system object. For recursive models, *all* coefficients are estimated recursively; you cannot mix fixed coefficients and recursive coefficients in your specification.

#### State equations

The state equations in recursive models must all be preceded by `@rcoef`. Follow `@rcoef` with the name of each state variable in the observation equations. You should list only one name after `@rcoef`, and you should have as many lines of `@rcoef` statements as there are state variables in your model.

*Example*

```
cs=b1+b2*gdp
m1=b3+b2*gdp
@rcoef b1
@rcoef b2
@rcoef b3
```

This recursive specification restricts the coefficient of GDP to be the same in the two observation equations.

**ARMA models***Observation equations*

The observation equations in ARMA models must all be preceded by the keyword `@identity`. Specify each equation after the keyword `@identity`, using a formula as you would in a system. The coefficients of the exogenous regressors and the MA parameters should be specified as elements of a coefficient vector.

The MA terms are specified in the observation equations using any lag of the state variables. You can skip any intervening MA terms. Each observation equation should have MA terms for exactly one state variable.

Since each observation equation has one and only one state variable, the number of observation equations and the number of state equations for ARMA models must be the same. This means that the MA coefficient matrix in the observation equation is restricted to be diagonal. As we will see below, there is no corresponding restriction on the AR coefficient matrix.

*State equations*

The state equations in ARMA models must all be preceded by the identifier `@state`. Each state equation specifies the AR terms of the equation with the corresponding state variable. `@state` in each state equation must be followed by the name of the state variable, an equals sign `=`, and the AR specification where the AR parameters should be elements of a coefficient vector. You can skip any intervening AR terms.

*Example*

```
@identity cs=c(1)+c(2)*gdp+su+c(3)*su(-4)
@identity m1=c(4)+c(5)*gdp+sv+c(6)*sv(-1)
@state su=c(7)*su(-1)+c(8)*sv(-1)
@state sv=c(9)*sv(-4)
```

The CS equation has an ARMA(1,4) error specification with the coefficients on MA(1), MA(2), MA(3) restricted to be zero, while the M1 equation has an ARMA(4,1) error specification with

the AR(1), AR(2), AR(3) coefficients restricted to be zero. The first state equation includes lags of both state variables, introducing correlation between the two observation equations.

### Time-Varying Parameter models

#### *Observation equations*

Each observation equation should be specified by formula, as in a system of equations. The observation equations may contain regressors with fixed coefficients and with random (time-varying) coefficients. Fixed coefficients should be specified as elements of coefficient vectors, while time-varying coefficients should involve a state variable.

To estimate the mean of the time-varying parameter, include a fixed coefficient in addition to the state variable in the observation equation. Otherwise, the time-varying parameter will be estimated with a zero mean.

#### *State equations*

The state equations must all be preceded by the keyword `@state`. Each state equation specifies the autoregressive dynamics of a time varying parameter; only random walk, AR(1), and constant specifications are allowed. To estimate the AR(1) parameter, specify the AR(1) coefficient as an element of a coefficient vector. For example,

```
@state sv1 = c(5)*sv1(-1)
```

You may impose a value for the AR(1) parameter. Consider the following examples:

```
@state sv = sv(-1)
@state su = .5*su(-1)
@state st = 0
```

The first equation specifies a random walk for the time-varying parameter SV, while the latter two equations restrict the AR(1) parameter to be .5 and 0, respectively. This latter specification is simply a pure random coefficients model.

If you wish to include a constant in the AR(1) state equation, you should enter it in the appropriate place in the observation equation, not in the state equation. You can either enter this term as a constant value, or as a coefficient, depending upon whether you wish to restrict or estimate the constant in the AR(1).

*Example*

```

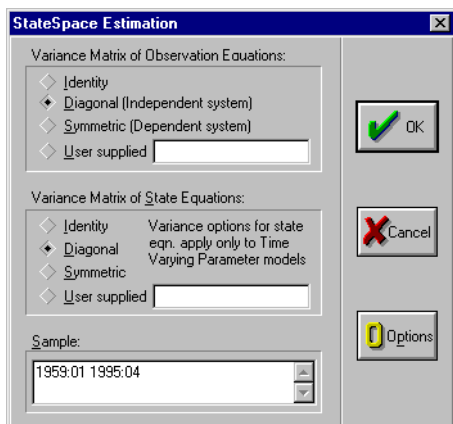
cs = c(1) + (su + c(2)) * gdp
m1 = c(3) + (sv + c(4)) * gdp
@state su = c(5) * su(-1)
@state sv = sv(-1)

```

Both observation equations have a fixed constant and a time-varying slope coefficient on GDP. The slope coefficient in the CS equation evolves as an unrestricted AR(1), while the slope coefficient in the M1 equation evolves as a random walk with drift.

## Estimating State Space Models

Once you have specified a state space model, you are ready to estimate the model; click **Estimate** on the toolbar or select **Procs/Estimate...**



You will need to specify the form of the covariance matrix of the error terms in both the observation and state equations:

- **Symmetric** allows the error terms to be contemporaneously correlated.

- **Diagonal** assumes the error terms are uncorrelated across equations. The off-diagonal terms are restricted to zero.

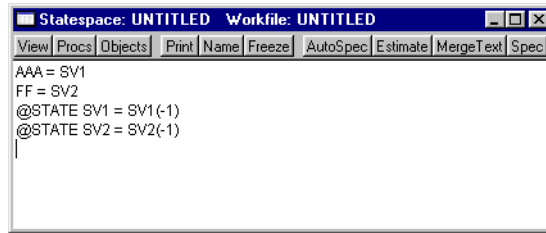
- **Identity** assumes the error terms are uncorrelated and have unit variance. The covariance matrix is restricted to be an identity matrix.

- **User supplied** allows you to specify the covariance matrix without estimating it. The covariance matrix must be stored as a named matrix object in EViews. Simply type the name of the matrix object to use for the covariance matrix.

Note that the choice between **Diagonal** and **Symmetric** options matters only for multiple equation models. For recursive estimation, the state variance is restricted to zero, while for ARMA estimation, the observation variance is restricted to zero. If you are estimating one of these latter two models, EViews will indicate that some variance options are not applicable by graying out the entries.

### Interpreting the estimation results

After you choose the variance options and click **OK**, EViews presents the estimation results in the state space window. For example, the specification:



when estimated, will display the state space output view. The top of this window contains equation results:

StateSpace: TVP\_D  
 Estimation Method: Maximum Likelihood  
 Date: 07/26/97 Time: 14:45  
 Model: Time-Varying Coefficient Model  
 Sample: 1959:01 1995:04  
 Included Observations: 434  
 Variance of observation equations: Diagonal  
 Variance of state equations: Diagonal  
 Convergence achieved after 32 iterations

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	2.362327	0.282745	8.354975	0.0000
C(2)	3.014641	0.087644	34.39650	0.0000
ObVar(1,1)	0.034207	0.004931	6.936846	0.0000
ObVar(2,2)	0.001758	0.066772	0.026334	0.9790
SSVar(1,1)	1.50E-07	972.4812	1.54E-10	1.0000
SSVar(2,2)	4.18E-06	37.29488	1.12E-07	1.0000
Final SV1	0.004943	0.000415	11.89977	0.0000
Final SV2	0.002640	0.002045	1.291211	0.1970
Log Likelihood	-578.2666			

The output header indicates that this is a time-varying coefficient model and that the covariance matrices of the observation and state equations were restricted to be diagonal. The top part of the table displays the estimated parameters and its standard errors. “ObVar” and “SSVar” are the elements of the estimated covariance matrix of the observation and state equations. The reported log likelihood is the maximized value of the multivariate normal log likelihood function.

The bottom part of the estimation output presents regression statistics for each equation:

AAA = C(1) +SV1*M1			
SV1 = SV1(-1)			
<hr/>			
R-squared	0.998123	Mean dependent var	8.117880
Adjusted R-squared	0.998123	S.D. dependent var	2.721182
S.E. of regression	0.117878	Sum squared resid	6.016638
Durbin-Watson stat	1.217322		
<hr/>			
FF = C(2) +SV2*M1			
SV2 = SV2(-1)			
<hr/>			
R-squared	0.999998	Mean dependent var	6.687558
Adjusted R-squared	0.999998	S.D. dependent var	3.425526
S.E. of regression	0.004814	Sum squared resid	0.010034
Durbin-Watson stat	2.569472		
<hr/>			

The regression statistics (e.g. R-squared) for time-varying and recursive coefficient models are computed using the smoothed residuals

$$y_t - A'x_t - H'\xi_{t|T} \tag{21.11}$$

from the observation equations. For ARMA models, the regression statistics use the smoothed residuals

$$\xi_{t+1|T} - F\xi_{t|T} \tag{21.12}$$

from the state equations.

Other estimation options

Initial values

As explained above, the estimation dialog provides options to choose the covariance matrices for the observation and state equations. You can also specify starting values for the state variable and its variance.

To set the starting values for the state variable and its variance, you must first create a named vector and matrix to hold the initial values.

For example, suppose you have a two equation state space object named SS1 and you want to set the initial values of the state vector to

$$\begin{bmatrix} SV1 \\ SV2 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \text{var} \begin{bmatrix} SV1 \\ SV2 \end{bmatrix} = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 2 \end{bmatrix}. \tag{21.13}$$

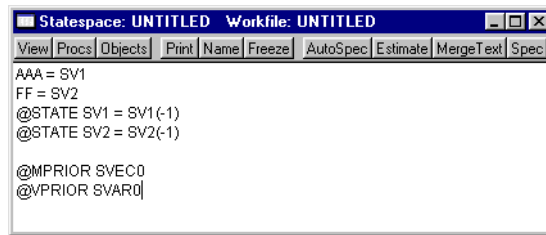
First create a named vector object, say SVEC0, to hold the initial values. Click **Objects/New Object**, choose **Matrix-Vector-Coeff** and enter the name SVEC0. Click **OK**, and then choose the type **Vector** and specify the size of the vector (in this case 2 rows). When you click **OK**, EViews will display the spreadsheet view of the vector SVEC0. Click the **Edit** +/- button to toggle edit

mode and type in the desired values. You can create a named matrix object, say SVAR0, in an analogous fashion.

Alternatively, you may find it easier to create and initialize the vector and matrix using commands. You can enter the following commands in the command window:

```
vector(2) svec0=0
matrix(2,2) svar0
svar0.fill(b=c) 1 0.5 0.5 2
```

Next, you must edit the specification window and include a line with the keywords `@mprior` and `@vprior` followed by the name of the object that contains the initial values. If you choose to set initial values, you must specify values for both the state variable and its variance:



Alternatively, type the following commands in the command window:

```
ssl.add @mprior svec0
ssl.add @vprior svar0
```

### Iteration control

EViews estimates state space models using numeric first derivative methods. This algorithm may converge slowly or even fail to achieve convergence. In such cases, you may want to modify the iteration algorithm by clicking the **Options** button in the estimation dialog. See Appendix E for a discussion of these options

## Working With the State Space

In this section we describe views and procedures that are specific to state space models; see Chapter 14 and Chapter 15 for an explanation of views and procedures common to all objects containing equations. Most of these views and procedures are not available until the state space specification has successfully been estimated.

### Views

The **Residual Correlation Matrix**, **Residual Covariance Matrix**, **Residual Graphs** all use the residuals from the observation equations for time-varying and recursive coefficient models and from the state equations for ARMA models.

**One-step ahead State Series Graphs** plot the one-step ahead conditional forecasts of the state series,  $\xi_{t+1|t}$ .

**Smoothed State Series Graphs** plot the smoothed state series,  $\xi_{t|T}$ .

## Procedures

**Make One-step ahead State Series and MSE** saves the one-step ahead conditional forecasts  $\xi_{t+1|t}$  and their standard errors  $s_{t+1|t}$  as named series in the current workfile. The one-step ahead conditional forecasts  $\xi_{t+1|t}$  and their mean squared errors  $P_{t+1|t}$  are computed as

$$\begin{aligned}\hat{\xi}_{t+1|t} &= F\xi_{t|t} \\ P_{t+1|t} &= FP_{t|t}F' + Q\end{aligned}\quad (21.14)$$

The standard errors  $s_{t+1|t}$  are the square roots of the diagonal elements of  $P_{t+1|t}$ .

**Make Smoothed State Series and MSE** saves the smoothed state series  $\xi_{t|T}$  and its standard errors  $s_{t|T}$  as named series in the current workfile. The smoothed series is the best estimate (in terms of mean squared error) of the state series using the full sample  $T$ . To compute the smoothed series, we first compute  $\xi_{t|t}$ ,  $\xi_{t|t-1}$ ,  $P_{t|t}$ , and  $P_{t|t-1}$  for  $t = 1, \dots, T$ . The smoothed series and its mean squared errors are then computed by the backward recursion,

$$\begin{aligned}\xi_{t|T} &= \xi_{t|t} + J_t(\xi_{t+1|T} - \xi_{t+1|t}) \\ P_{t|T} &= P_{t|t} + J_t(P_{t+1|T} - \xi_{t+1|t})J_t'\end{aligned}\quad (21.15)$$

for  $t = T-1, T-2, \dots, 1$ , where  $J_t = P_{t|t}F'P_{t+1|t}^{-1}$ . The standard errors  $s_{t|T}$  are the square roots of the diagonal elements of  $P_{t|T}$ .

**Make Residuals** saves the residuals as a named series in the current workfile. The residuals are computed from the observation equations as

$$y_t - A'x_t - H'\xi_{t|t-1}, \quad (21.16)$$

for time-varying and recursive coefficient models, and from the state equations as

$$\xi_{t+1|t} - F\xi_{t|t-1}, \quad (21.17)$$

for ARMA models. Note that these residuals are *not* the smoothed residuals used to compute the regression statistics.

**Make Standardized Residuals Series** saves the standardized residuals as a named series in the current workfile. The standardized residuals are computed as

$$(H'P_{t|t-1}H + R)^{-1/2}(y_t - A'x_t - H'\xi_{t|t-1}) \quad (21.18)$$

for time-varying and recursive coefficient models, and as

$$(H'P_{t|t-1}H + R)^{-1/2}(\xi_{t+1|t} - F\xi_{t|t-1}) \quad (21.19)$$



for ARMA models. The standardized residuals should have variance (close to) one.

**Make Model** creates a model object containing the state space specification and estimation results.

## Regression Statistics

All of the regression statistics for systems described in Appendix A are available for state space objects. You may access individual coefficients,  $t$ -statistics, and  $p$ -values, as well as summary statistics for each observation equation.

## Simulation Example

As an illustration of working with a state space model, we simulate, then estimate, a model consisting of a random walk plus noise:

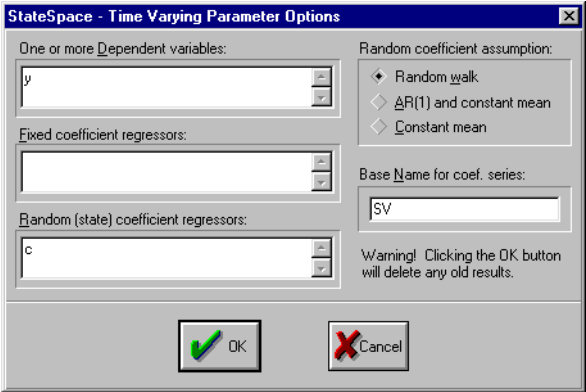
$$\begin{aligned} y_t &= \xi_t + w_t \\ \xi_t &= \xi_{t-1} + v_t \end{aligned} \tag{21.20}$$

where  $w$  and  $v$  are random draws from the normal distribution. This model fits the framework of a time-varying parameter model, where  $y$  is the dependent variable and  $\xi_t$  is a random coefficient on the constant term. We can simulate the series by the commands:

```
' create quarterly workfile
create q 70 90
' generate state series
smpl @first @first
series ss=0
smpl @first+1 @last
series ss=ss(-1)+nrnd
' generate dependent variable
series y=ss+2*nrnd
```

Note that the variance of the state equation is 1, while the variance of the observation equation is 4.

To estimate a time varying parameter model, first create a state space object: click **Objects/New Object...**, select **SSpace**, and click **OK**. Then click **AutoSpec/Time-Varying Parameter** and fill in the dialog as follows:



When you click OK, EViews displays the specification in the state space window as follows:

Y = SV1

@STATE SV1 = SV1(-1)

To estimate the model, click **Estimate** and then click **OK**, leaving the variance options at the default **Diagonal** setting. When the estimation procedure converges, EViews will display the estimation results in the state space window:

SSpace: UNTITLED  
Estimation Method: Maximum Likelihood  
Date: 11/19/97 Time: 13:  
Model: Time-Varying Coefficient Model  
Sample: 1970:1 1990:4  
Included Observations: 83  
Variance of observation equations: Diagonal  
Variance of state equations: Diagonal  
Convergence achieved after 21 iterations

	Coefficient	Std. Error	t-Statistic	Prob.
ObVar(1,1)	3.877500	0.919621	4.216411	0.0001
SSVar(1,1)	1.192663	0.588023	2.028258	0.0459
Final SV1	-0.420919	1.681654	-0.250301	0.8030
Log Likelihood	-201.2521			
Y = SV1				
SV1 = SV1(-1)				
-----				
R-squared	0.729401	Mean dependent var	-0.812767	
Adjusted R-squared	0.732661	S.D. dependent var	3.247657	
S.E. of regression	1.679196	Sum squared resid	234.0350	
Durbin-Watson stat	2.891476			

The estimated variances of the observation and state equations are fairly close to the known “true” values of 4 and 1, respectively.

We can compare the actual values of the state series to those obtained from the estimation to see how well the estimation recovered our simulated random walk. The smoothed state series provides our best estimate of the state series using all of the information in the sample. To obtain the smoothed state series, click **Procs/Make Smoothed State Series and MSE**. EViews will display the smoothed state series SV1SM and its standard errors SV1SM\_SE in an untitled group window.

Once created, we can plot the smoothed state series SV1SM together with the actual state series SS with the command

```
plot ss sv1sm
```

The estimation appears to have performed fairly well at recovering the random walk from the noisy data.

We can examine how increasing the “noise” in the observation equation causes the estimates of the state series to deteriorate, by changing the variance of the observation equation. For example, we can increase the variance of the observation equation by generating the dependent variable using the command:

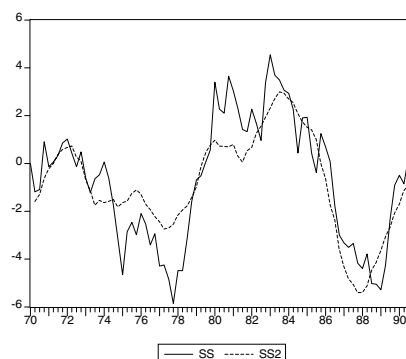
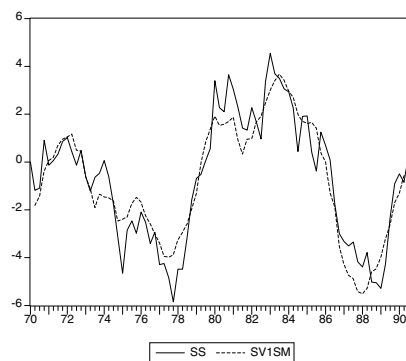
```
series y=ss+4*nrnd
```

The observation now has a variance of 16. We can follow the same procedure above to obtain the smoothed state series and plot it against the actual state series:

Comparing this plot to the previous one, we see that our estimates of the state vector are not as good as for the earlier example. The increased noise makes it harder to extract the true random walk from the observed data.

## Commands

To create a new state space object, follow the `sspace` command with a name for the state space object:



```
sspace tvp
```

creates and opens the window for the new state space object named TVP.

To estimate a state space object, follow the name of the state space with the word `ml`:

```
tvp1.ml
```

To store the smoothed state series as a named object, use the `makessmooth` procedure:

```
tvp1.makessmooth smooth1 smooth1_se
```

stores the smoothed series and standard errors of the first state variable in TVP1 as series named `SMOOTH1` and `SMOOTH1_SE`, respectively.

See the *Command and Programming Reference* for a complete list of commands and options available for state space objects.

## Technical Discussion

### Iterative algorithm

To start the algorithm, we need initial values for the state vector  $\xi_0$ , its variance  $P_0$ , and the system parameters  $\beta^{(0)} = \{A_0, H_0, F_0, R_0, Q_0\}$ . (We describe below how we obtain these initial values). Given these initial values, we first compute the sequence of state series  $\xi$  and its mean squared errors  $P$  using the Kalman filter. These are then used to evaluate the Gaussian log likelihood function and to find the next set of parameters  $\beta^{(1)}$ . The iterative methods for finding  $\beta^{(i+1)}$  from  $\beta^{(i)}$  are described in Appendix E. This completes one iteration. The next iteration uses the initial values  $\xi_0$ ,  $P_0$ , and  $\beta^{(1)}$  to obtain  $\beta^{(2)}$  and so on, until convergence is achieved.

### Initial values

There are two sets of starting values for state space models; one for the state vector  $\xi_0$  and its variance and one for the coefficients. By default, EViews estimates these initial values by running OLS on the observation equation treating  $\xi_0$  as a fixed parameter. This OLS regression uses the first  $k$  observations, where  $k$  is the number of parameters in the observation equation.

EViews provides options to specify starting values other than the default. To set starting values for the coefficients, use the `param` command to set the values in the C coefficient vector. For example, to set the initial values of the first two parameters to 0.1, type

```
param c(1) 0.1 c(2) 0.1
```

in the command window.

To set starting values for the state vector  $\xi_0$  and its variance  $P_0$ , you must first save the starting values as a named vector (for  $\xi_0$ ) and a named matrix (for  $P_0$ ) in the workfile. Then edit the specification view of the state space object and add a line with the command `@mprior` (for  $\xi_0$ ) and `@vprior` (for  $P_0$ ) followed by the names of the vector and matrix object. For example, to set the initial values  $\xi_0 = [0, 0]'$  and  $P_0 = I_2$ , type

```
vector(2) svec0=0
matrix(2,2) svar0=@identity(2)
ss1.add @mprior svec0
ss1.add @vprior svar0
```

Note: to specify starting values for the state vector, you must specify *both* the mean  $\xi_0$  and its variance  $P_0$ . Note also that the initial values for the variance matrix  $P_0$  should be a symmetric positive definite matrix.

### Error messages

Sometimes you may experience difficulties in estimating a state space model. You should try other starting values (as explained above) or modify the iteration algorithm using **Options** in the estimation dialog. Be aware that identification problems in state space models are quite common; see Hamilton (1994a, 387–388).



## Chapter 22. Pooled Time Series, Cross-Section Data

---

Data often contain information on cross-sectional units observed over time. In many cases, a relatively small number of cross-sectional units are observed over a number of periods. For example, you may have time series data on GDP for a number of European nations. Or perhaps you have state level data on unemployment observed over time. We term such data *pooled time series*, *cross-section data*.

EViews provides a number of specialized tools to help you work with pooled data. EViews will help you manage your data, perform operations in either the time series or the cross-section dimension, and apply estimation methods that account for the pooled structure of your data.

The EViews object that manages time series/cross-section data is called a *pool*. The remainder of this chapter will describe how to set up your data to work with pools, and how to define and work with pool objects.

### Creating a Workfile For Pooled Data

Your first step will be to create a workfile to hold your pooled data. There is nothing special about the creation of this workfile. Simply select **File/New/Workfile...** and enter the frequency and dates for your workfile.

The range of your workfile should represent the earliest and latest dates that you wish to consider for any of the cross-section units. For example, if you want to work with data for Japan from 1950 to 1993, and data for the United States from 1956 to 1997, you should create a workfile ranging from 1950 to 1997.

### The Pool Object

At the heart of a pool object is a list of names that you will use to refer to cross-section members. For reasons that will become apparent, these names should be relatively short. For example, in a cross-country study, you might use “USA” to refer to the United States, “CAN” to refer to Canada, “KOR” to identify Korea, “JPN” for Japan, and “UK” for the United Kingdom.

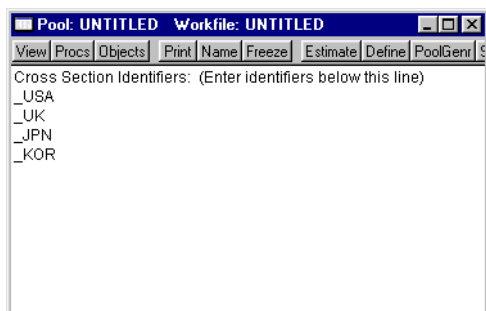
Defining the cross-section identifiers in a pool tells EViews about the structure of your data. In the example above, EViews will understand that when using this pool, you wish to work with separate time series data for each of the countries.

It is important to note that the pool object does not itself contain series or data. A pool object is simply a description of the underlying structure of your data. Consequently,

deleting the pool will not delete the series referenced by the pool, and modifying the series referenced by the pool will modify the data used by the pool.

## Creating a Pool Object

To create a pool object, select **Objects/New Object/Pool...** and enter the identifiers in the edit window.



There are no special restrictions on the labels that you can use for cross-section identifiers, though you must be able to form legal EViews series names containing these identifiers. Here we have prepended the “\_” character to each of the identifiers; this is not necessary, but we find that it makes it easier to spot the identifier when it is used as part of a series name.

## Viewing or Editing a Pool Definition

To view the cross-section identifiers in your pool, push the **Define** button on the toolbar, or select **View/Cross-Section Identifiers**. If desired, you can edit the list of identifiers.

## Copying a Pool Object

Typically, you will work with more than one pool object. Multiple pools are used to define various subsamples of cross-section identifiers, or to work with different pooled estimation specifications.

To copy a pool object, open the original pool, and select **Objects/Copy Object...** Alternatively, you can highlight the name of the pool in the workfile window, select **Objects/Copy Selected...** and enter the new name.

## Working with Pools and Series

Even though you will be working with pooled data, all of your data will be held in ordinary EViews series. These series can be used in the usual ways: among other things, they can be tabulated, graphed, used to generate new series, or used in estimation. You can also use the pool object to work with the individual series.

## Naming Your Series

The key to using series with pools is to name your series using a combination of a *base name* and the cross-section identifier. The cross-section identifier may be embedded at an arbitrary location in the series name, so long as this is done consistently across identifiers.



For example, suppose that you have a pool object that contains the identifiers `_JPN`, `_USA`, `_UK`, and that you have time series measurements on GDP for each of the cross-section units. We will use “GDP” as the base name for our series.

You can choose to place the identifier at the end of the base name, in which case, you should name your series `GDP_JPN`, `GDP_USA`, and `GDP_UK`. Alternatively, you can elect to put the section identifiers, `JPN_`, `USA_`, `UK_`, in front of the name, so the series will be named `JPN_GDP`, `USA_GDP`, `UK_GDP`. The identifiers may also be placed in the middle of series names—for example, `GDP_JPNINFO`, `GDP_USAINFO`, `GDP_UKINFO`.

It really doesn’t matter whether the identifiers are used at the beginning, end, or middle of your names; you should pick the naming style that you find easiest to work with. Consistency, however, is absolutely essential. You should not name your three series `JPNGDP`, `GDPUSA`, `UKGDP1`, as EViews would be unable to reference these series using a pool object.

## Pool Series

Once the series names have been chosen to correspond with the identifiers in your pool, the pool object can be used to work with a set of series. The key to this processing is the concept of a *pool series*.

A pool series is actually a group of series defined by a base name and the entire list of cross-section identifiers. Pool series are specified using the base name, and a “?” character placeholder for the cross-section identifier. If your series are named `GDPJPN`, `GDPUSA`, `GDPUK`, the corresponding pool series will be referred to as `GDP?`. If the names of your series are `JPNGDP`, `USAGDP`, and `UKGDP`, the pool series will be `?GDP`.

When you use a pool series name, EViews understands that you wish to work with all of the series in the pool series. EViews loops through the list of cross-section identifiers and substitutes each identifier in place of the “?”. EViews will process your instructions using the constructed names.

A pool series may only be accessed through the pool object, since the placeholder “?” has no meaning without a list of cross-section identifiers. If you attempt to use a pool variable outside of a pool object, you will receive an error message saying that your variable is not defined, or EViews will interpret the “?” as a wildcard character (see Appendix D).

## Importing Pooled Data

There are several ways to import data so that they may be used in pooled data analysis. Before considering the various methods, we will need to understand the structure of pooled time series, cross-section data, and to distinguish between data that are in *unstacked* and *stacked* form.

A given piece of information in a pooled setting may be indexed along three dimensions: the variable, the time period, and the cross-section. For example, you may be interested in the value of GDP, in 1989, for Korea.

Working with your data in three dimensions is difficult, so typically you will employ a two-dimensional representation of your data. There is no unique way to organize three-dimensional data in two-dimensions, but several formats are commonly employed.

### Unstacked Data

In this form, observations on a given variable for a given cross-section are grouped together, but are separated from observations for other variables and other cross sections. For example, suppose the top of our data file contains the following:

year	c_usa	c_kor	c_jpn	g_usa	g_jpn	g_kor
1954	61.6	77.4	66	17.8	18.7	17.6
1955	61.1	79.2	65.7	15.8	17.1	16.9
1956	61.7	80.2	66.1	15.7	15.9	17.5
1957	62.4	78.6	65.5	16.3	14.8	16.3
⋮	⋮	⋮	⋮	⋮	⋮	⋮

Here, the base name C represents consumption, while G represents government expenditure. Each country has its own separately identified column for consumption, and its own column for government expenditure.

EViews works naturally with data that are *unstacked* and will read unstacked data using the standard input procedures described in Chapter 4, “Basic Data Handling”, page 62. Simply read each cross-section specific variable as an individual series, making certain that the names of the series follow the pool naming conventions described above.

### Stacked Data

Pooled data can also be arranged in stacked form, where all of the data for a variable are grouped together, but separated from data for other variables. In the most common form, the data for different cross-sections are stacked on top of one another, with each column representing a variable:

---

id	year	c	g
_usa	1954	61.6	17.8
⋮	⋮	⋮	⋮
_usa	1992	68.1	13.2
_uk	1954	62.4	23.8
⋮	⋮	⋮	⋮
_uk	1992	67.9	17.3
_jpn	1954	66	18.7
⋮	⋮	⋮	⋮
_jpn	1992	54.2	7.6
_kor	1954	77.4	17.6
⋮	⋮	⋮	⋮
_kor	1992	na	na

---

We say that these data are *stacked by cross-section*. Alternatively, we may have data that are *stacked by date*:

per	id	c	g
1954	_usa	61.6	17.8
1954	_uk	62.4	23.8
1954	_jpn	66	18.7
1954	_kor	77.4	17.6
⋮	⋮	⋮	⋮
1992	_usa	68.1	13.2
1992	_uk	67.9	17.3
1992	_jpn	54.2	7.6
1992	_kor	na	na

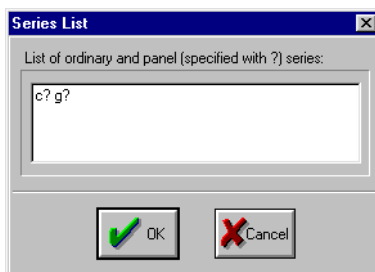
Each column again represents a variable, but within each column, the data are arranged by year. If data are stacked by year, you should make certain that the ordering of the cross-sectional identifiers within a year is consistent across years.

### Manual Entry/Cut-and-Paste

One way to import data is by manually entering, or copying-and-pasting into a stacked representation of your pool series:

- First, specify which time series observations will be included in your stacked spreadsheet by setting the workfile sample.
- Next, open the pool, then select **View/Spreadsheet View...** EViews will prompt you for a list of series. You can enter ordinary series names or pool series names. If the series exist, then EViews will display the data in the series. If the series do not exist, then EViews will create the series or group of series, using the cross-section identifiers if you specify a pool series.
- EViews will open the stacked spreadsheet view of the pool series. If desired, click on the **Order** +/- button to toggle between stacking by cross-section and stacking by date.
- Click **Edit** +/- to turn on edit mode in the spreadsheet window, and enter your data, or cut-and-paste from another application.

For example, if we have a pool that contains the identifiers `_USA`, `_UK`, `_JPN`, and `_KOR`, we can instruct EViews to create the series `C_USA`, `C_UK`, `C_JPN`, `C_KOR`, and `G_USA`, `G_UK`, `G_JPN`, `G_KOR` by entering:



EViews will open a stacked spreadsheet view of the newly created series. Here we see the series stacked by cross-section, with the pool series names in the column header, and the cross-section/date identifiers labeling each row:

obs	C?	G?
_USA-1950	NA	NA
_USA-1951	NA	NA
_USA-1952	NA	NA
_USA-1953	NA	NA
_USA-1954	NA	NA
_USA-1955	NA	NA
_USA-1956	NA	NA
_USA-1957	NA	NA
_USA-1958	NA	NA
_USA-1959	NA	NA
_USA-1960	NA	NA
_USA-1961	NA	NA
_USA-1962	NA	NA
_USA-1963	NA	NA

If desired, click on **Order** +/- to toggle between stacking methods. Click on **Edit** +/- to turn on edit mode, and enter or cut-and-paste into the window.

## File Import

You can import stacked data from a file into individual series using a pool object. While the data in the file may be stacked either by cross-section or by period, EViews requires that:

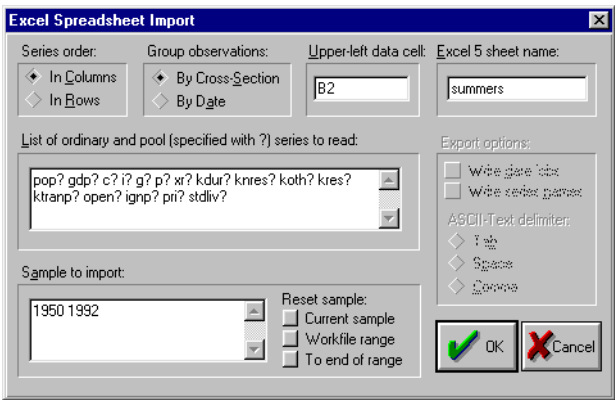
- The stacked data are *balanced*.
- The cross-sections are ordered in the file in the same way that the cross-sectional identifiers are listed in the pool.

By “balanced”, we mean that if the data are stacked by cross-section, each cross-section should contain exactly the same number of periods—if the data are stacked by date, each date should have exactly the same number of cross-sectional observations arranged in the same order.

We emphasize that *the underlying data need not be balanced, only the representation in the import file*. If you have missing values for some observations, you should make certain that there are lines in the file representing the missing values. In the two examples above, the underlying data are not balanced, since information is not available for Korea in 1992. The data in the file have been balanced by including an observation for the missing data.

To read data from a file using a pool object, first open the pool, then select **Procs/Import Pool data (ASCII, .XLS, .WK?)...** It is important that you use the import procedure associated with the pool object, and not the standard file import procedure.

Select your input file in the usual fashion. If you select a spreadsheet file, EViews will open a spreadsheet import dialog prompting you for additional input.



- Indicate whether the pool series are in rows or in columns, and whether the data are stacked by cross-section, or stacked by date.
- In the edit box, enter the names of the series you wish to import. These series names can be ordinary series names or pool names.
- Fill in the sample information, starting cell location, and optionally, the sheet name.

When you specify your series using pool series names, EViews will, if necessary, create and name the series using the cross-section identifiers. If you list an ordinary series name, EViews will, if needed, create the single series.

EViews will read from your file into the specified variables using the sample information. If you enter an ordinary series name, EViews will read multiple values into the series, so that, upon completion, the series will contain the last set of values read from the file.

The basics of importing from ASCII files are analogous, but the corresponding dialog contains many additional options to handle the complexity of ASCII files. For details, see “Addendum: Reading ASCII Files” on page 70.

## Exporting Pooled Data

You can export your data by reversing the procedures described above for data input. Since EViews allows you to read and write data that are unstacked, stacked by cross-section, or stacked by date, you can use EViews to restructure your data in accordance with your needs.

## Working with Pooled Data

The underlying series for each cross-section member are ordinary series, so all of the EViews tools for working with the individual cross-section series are available. In addition, EViews provides you with a number of specialized tools which allow you to work with your pool data. Using EViews, you can perform, in a single step, similar operations on all the series corresponding to a particular pooled variable.

### Examining Data

As described above, you can view your data in stacked spreadsheet form. Select **View/Spreadsheet View...**, and list the series you wish to display. The names can include both ordinary and pool series names. Click on the **Order** +/- button to toggle between stacking your observations by cross-section and by date.

We emphasize that stacking your data only provides an alternative view of the data, and does not change the structure of the individual series in your workfile. Stacking data is not necessary for any of the data management or estimation procedures described below.

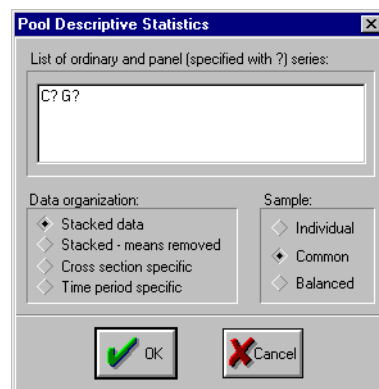
### Describing Data

You may compute descriptive statistics for your series using a pool object. Select **View/Descriptive Statistics...** from the pool toolbar.

In the edit box, you should list the ordinary and pooled series for which you want to compute the descriptive statistics. EViews will compute the mean, median, minimum, maximum, standard deviation, skewness, kurtosis, and the Jarque-Bera statistic for these series.

Next you should choose between the three sample options:

- **Individual:** uses the maximum number of observations available. If an observation on a variable is available for a particular cross-section, it is used in computation.



- **Common:** uses an observation only if data on the variable are available for all cross-sections in the same period. This method is equivalent to performing listwise exclusion by variable, then cross-sectional casewise exclusion within each variable.
- **Balanced:** includes observations when data on all variables in the list are available for all cross-sections in the same period. The balanced option performs casewise exclusion by both variable and cross-section.

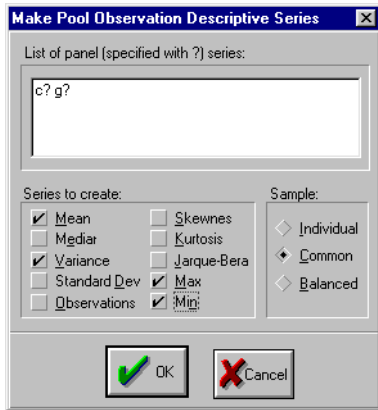
Lastly, you should choose the computational method corresponding to one of the four data structures:

- **Stacked data:** display statistics for each variable in the list, computed over all cross-sections and periods. These are the descriptive statistics that you would get if you ignored the pooled nature of the data, stacked the data, and computed descriptive statistics.
- **Stacked – means removed:** compute statistics for each variable in the list after removing the cross-sectional means, taken over all cross-sections and periods.
- **Cross-section specific:** show the descriptive statistics for each cross-sectional variable, computed across all periods. These are the descriptive statistics derived by computing statistics for the individual series.
- **Time period specific:** compute period-specific statistics. For each period, compute the statistic using data on the variable from all the cross-sectional units in the pool.

Be aware that the latter two methods may generate a great deal of output. Cross-section specific computation generates a set of statistics for each variable/cross-section combination. If you ask for statistics for three pool variables and there are 20 cross-sections in your pool, EViews will compute descriptive statistics for 60 series. For time-period specific computation you will compute a set of statistics for each date/variable combination. If you have a sample with 100 periods and you list three pool variables, EViews will compute 300 sets of statistics.

You can save the period-specific statistics in series objects. Select **Procs/Make Period Stat Series...** from the pool window, and fill out the dialog.





In the edit window, list the series for which you wish to calculate period-statistics. Next, select the particular statistics you wish to compute, and choose a sample option.

EViews will save your statistics in new series, and will open an untitled group window to display the results. The series will be named automatically using the base name followed by the name of the statistic (MEAN, MED, VAR, SD, OBS, SKEW, KURT, JARQ, MAX, MIN). In this example, EViews will save the statistics using the names CMEAN, GMEAN, CVAR, GVAR, CMAX, GMAX, CMIN, and GMIN.

## Generating Data

You can generate or modify pool series using the **PoolGenr** procedure. Click on **PoolGenr** on the pool toolbar and enter your formula as you would for a regular **Genr**, using pool names as appropriate. Using our example from above, entering:

```
ratio? = g?/g_usa
```

is equivalent to entering the following four commands:

```
ratio_usa = g_usa/g_usa
ratio_uk  = g_uk/g_usa
ratio_jpn = g_jpn/g_usa
ratio_kor = g_kor/g_usa
```

**PoolGenr** applies the formula you supply using an implicit loop across cross-section identifiers, creating or modifying series as appropriate.

You may use **PoolGenr** and **Genr** together to generate new variables. For example, to create a dummy variable that is equal to 1 for the US and 0 for all other countries, first select **PoolGenr** and enter:

```
dum? = 0
```

to initialize all of the dummy variable series to 0. Then, to set the US values to 1, select **Quick/Generate Series...** from the main menu, and enter:

```
dum_usa = 1
```

To modify a set of series using a pool, select **PoolGenr**, and enter the new pool series expression:

```
dum? = dum? * (g? > c?)
```

You can also use the implicit looping feature to perform calculations across cross-sectional units in a given period. For example, if we have an ordinary series **SUM** which is initialized to zero, then a **PoolGenr** using the expression:

```
sum = sum + c?
```

will calculate the ordinary series given by the ordinary **Genr**:

```
sum = c_usa + c_uk + c_jpn + c_kor
```

Note that this example is merely to illustrate the notion of implicit looping. As we saw above, EViews provides you with built-in features to compute period-specific statistics of this type.

### Make Pool Group

You may wish to work with a set of pool series using standard EViews tools for group objects. Select **Procs/Make Group...** and enter the names of ordinary and pool series. EViews will create an untitled group object containing the series.

### Deleting/Storing/Fetching Data

Pools may be used to delete, store, or fetch sets of series. Simply select **Procs/Delete pool series...**, **Procs/Store pool series (DB)...**, or **Procs/Fetch pool series (DB)...** as appropriate, and enter the ordinary and pool series names of interest.

If, for example, you instruct EViews to delete the pool series **C?**, EViews will loop through all of the cross-section identifiers and delete all series whose names begin with the letter “C” and end with the cross-section identifier.

## Pooled Estimation

There are a number of ways that you can use information about the structure of your pooled data in estimating an equation. You might estimate a fixed or random intercept model, or perhaps a model with selected variables that have different coefficients across cross-sections, as well as separate AR(1) coefficients. Or you could estimate a separate equation for each cross-sectional unit.

EViews pool objects allow you to estimate your model using least squares, weighted least squares with estimated cross-section weights, or seemingly unrelated regressions, all without rearranging or reordering your data.

Below, we describe how you can use pools and systems to estimate more general and complex models, including two-stage least squares and nonlinear specifications, or models with complicated cross-section coefficient restrictions.

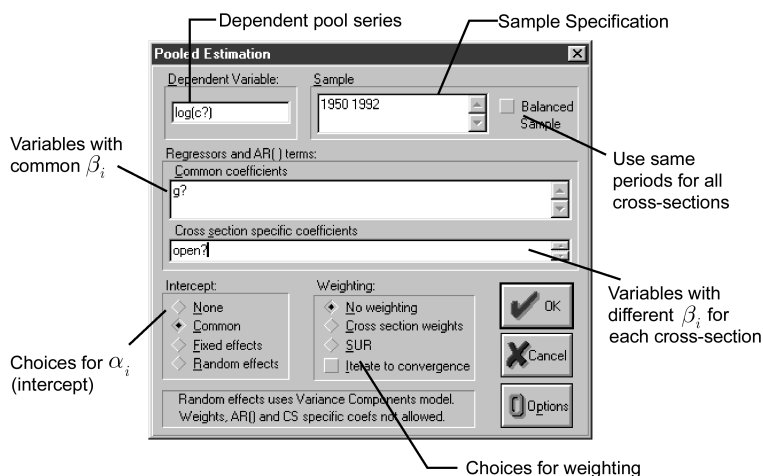
## How to Estimate a Pool Equation

Generally speaking, pool objects can be used to estimate equations of the form:

$$y_{it} = \alpha_{it} + x_{it}'\beta_i + \epsilon_{it}, \quad (22.1)$$

for  $i = 1, 2, \dots, N$  cross-section units and periods  $t = 1, 2, \dots, T$ . Further details and discussion are provided later in this chapter.

Press the **Estimate** button on your pool toolbar, and the following dialog will open:



### Dependent Variable

List a pool variable, or an EViews expression containing a pool variable, in the Dependent Variable edit box.

### Sample

Enter your sample specification in the edit window at the upper right.

By default, EViews will use the largest sample possible in each cross-section. An observation will be excluded if any of the explanatory or dependent variables *for that cross-section* are unavailable in that period.

The checkbox for **Balanced Sample** instructs EViews to perform listwise exclusion over all cross-sections. EViews will eliminate an observation if data are unavailable *for any cross-section* in that period. This exclusion ensures that estimates for each cross-section will be based on a common set of dates.

If all of the observations for a cross-section unit are not available, that unit will temporarily be removed from the pool for purposes of estimation. The EViews output will inform you if any cross-section units were dropped.

### Explanatory Variables

Next, you will list your regressors. There are two edit boxes where you will enter your explanatory variables:

**Common coefficients:** — enter variables that are to have the same coefficient across all cross-section members of the pool. EViews will include a single coefficient for each variable, and will label the output using the ordinary or pool name, as appropriate.

**Cross-section specific coefficients:** — list variables with different coefficients for each member of the pool. EViews will include a different coefficient for each cross-sectional unit, and will label the output using the cross-section identifier followed by “—” and then the ordinary series name.

For example, if you include the ordinary variable TIME and POP? in the common coefficient list, the output will include estimates for TIME and POP?. If you include these variables in the cross-section specific list, the output will include coefficients of the form \_USA—TIME, \_UK—TIME, and \_USA—POP\_USA, \_UK—POP\_UK, etc.

Be aware that estimating your model with cross-section specific variables may generate large numbers of coefficients—the number of these coefficients equals the product of the number of pool identifiers and the number of variables in the list.

You may include AR terms in your specification. If the terms are entered in the common coefficients list, EViews will estimate the model assuming a common AR error. If the AR terms are entered in the cross-section specific list, EViews will estimate separate AR terms for each member. See Chapter 13 for a description of AR specifications.

Note that EViews only allows specification by list for pool equations. If you wish to estimate a nonlinear specification, you must first create a system object, and then edit the system specification (see “Creating Systems Using Pools” on page 544).

### Intercept

In the area labeled **Intercept**: you can choose between alternative specifications for  $\alpha_i$ :

None	no intercepts; $\alpha_{it} = 0$ .
Common	identical intercept for all pool members: $\alpha_{it} = \alpha$ .

Fixed effects	different intercepts estimated for each pool member: $\alpha_{it} = \alpha_i$ .
Random effects	treats intercepts as random variables across pool members: $\alpha_{it} = \alpha_i$ , $E(\alpha_i \epsilon_{it}) = 0$ .

You cannot estimate random effects models with cross-section specific coefficients, AR terms, or weighting.

### Weights

EViews does not weight observations in pooled estimation by default, but you have the option of estimating weighted versions of your specifications. There are three options for weights:

No weighting	all observations are given equal weight.
Cross section weights	GLS using estimated cross-section residual variances.
SUR	analogue to Seemingly Unrelated Regression—GLS using estimated cross-section residual covariance matrix.

If you select **Cross section weights**, EViews will estimate a feasible GLS specification assuming the presence of cross-section heteroskedasticity. If you select **SUR**, EViews estimates a feasible GLS specification correcting for both cross-section heteroskedasticity and contemporaneous correlation. This specification is sometimes referred to as the Parks estimator. More detailed descriptions of these methods are provided in the “Technical Discussion” on page 545 below.

Bear in mind that there are a number of potential pitfalls associated with SUR/Parks estimation (see Beck and Katz (1995)). Furthermore, EViews may be unable to compute estimates for this model when you have large numbers of cross-sections or a small number of time periods. The average number of periods used in estimation must be at least as large as the number of cross-section units. Even if you have a sufficient number of observations, the estimated residual correlation matrix must also be nonsingular. If either condition is violated, EViews will display a “Near Singular Matrix” error.

There is a checkbox labeled **Iterate to convergence** which controls the feasible GLS procedures. If selected, EViews will continue to update the weights and coefficients until they converge. This option has no effect if your model includes AR terms since EViews will always iterate weights and coefficients to convergence in AR specifications.

## Options

### *Iteration and Convergence Options*

If you select weighted estimation and ask EViews to iterate to convergence, you can control the iterative process by specifying convergence criterion and the maximum number of iterations. Press the **Options** button in the equation dialog box and enter the desired values.

### *White Heteroskedasticity Covariance*

EViews can estimate covariances that are robust to general heteroskedasticity. This form of heteroskedasticity is more general than the cross-section heteroskedasticity described above, since variances within a cross-section are allowed to differ across time.

To request White standard errors and covariances, press the **Options** button, and select White Heteroskedasticity Consistent Covariance. Note that this option is not available with SUR and random effects estimation.

## Pool Equation Examples

As an illustration, we take data from the Penn World Table for GDP for seven countries: Canada, France, Germany, Italy, Japan, United Kingdom, and the US. The corresponding pool identifiers are \_CAN, \_FRA, \_GER, \_ITA, \_JPN, \_UK, \_US.

First, we estimate a model regressing GDP? on GDP?(-1) (note that the '?' precedes the lag specification). All coefficients are restricted to be the same across all cross-sections, so this is equivalent to estimating a model on the stacked data, ignoring all cross-sectional information:

The output from this regression is given by:

Dependent Variable: GDP?  
 Method: Pooled Least Squares  
 Date: 10/20/97 Time: 15:45  
 Sample(adjusted): 1951 1992  
 Included observations: 42 after adjusting endpoints  
 Total panel observations 294

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	245.3620	41.47131	5.916428	0.0000
GDP?(-1)	1.000448	0.004083	245.0305	0.0000
R-squared	0.995160	Mean dependent var	9681.432	
Adjusted R-squared	0.995144	S.D. dependent var	3786.743	
S.E. of regression	263.8918	Sum squared resid	20334560	
Log likelihood	-1818.911	F-statistic	60039.95	
Durbin-Watson stat	1.488062	Prob(F-statistic)	0.000000	

Note that there is a single coefficient for both the constant C, and the pool series for lagged GDP.

Next, suppose we estimate the same specification, but select **Fixed effects** for the intercept specification, and select **Cross-section weights** for our weighting. This implies that each pool will have an unrestricted intercept, and that each pool equation is downweighted by an estimate of the cross-section residual standard deviation. The results are given by:

Dependent Variable: GDP?  
 Method: GLS (Cross Section Weights)  
 Date: 10/20/97 Time: 15:49  
 Sample: 1951 1992  
 Included observations: 42  
 Total panel observations 294  
 Convergence achieved after 1 iteration(s)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
GDP?(-1)	1.004838	0.004088	245.8188	0.0000
Fixed Effects				
_CAN--C	189.8252			
_FRA--C	197.0985			
_GER--C	239.3465			
_ITA--C	208.5929			
_JPN--C	298.8015			
_UK--C	136.0450			
_US--C	158.0119			
Weighted Statistics				
R-squared	0.996134	Mean dependent var	10507.92	
Adjusted R-squared	0.996039	S.D. dependent var	4168.432	
S.E. of regression	262.3466	Sum squared resid	19684159	
Log likelihood	-1806.950	Durbin-Watson stat	1.504483	
Unweighted Statistics				
R-squared	0.995313	Mean dependent var	9681.432	
Adjusted R-squared	0.995199	S.D. dependent var	3786.743	
S.E. of regression	262.3922	Sum squared resid	19691008	
Durbin-Watson stat	1.543536			

Note that EViews displays both the estimates of the fixed effects, and summary statistics for the weighted and unweighted models. The estimates of the fixed effects do not have reported standard errors since EViews treats them as nuisance parameters for the purposes of estimation. If you wish to compute standard errors for the cross-section constants, you should estimate a model without a constant and explicitly enter the C in the **Cross section specific coefficients** edit field.

We can generalize this specification by estimating a separate coefficient for lagged GDP in each cross section. Simply click on the **Estimate** button to reestimate the pool, move GDP?(-1) to the **Cross section specific coefficients** edit field, and click OK. The top portion of the output is given below:

Dependent Variable: GDP?  
Method: GLS (Cross Section Weights)  
Date: 10/20/97 Time: 15:56  
Sample: 1951 1992  
Included observations: 42  
Total panel observations 294  
Convergence achieved after 1 iteration(s)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
_CAN--GDP_CAN(-1)	0.998332	0.016368	60.99230	0.0000
_FRA--GDP_FRA(-1)	0.997262	0.008985	110.9879	0.0000
_GER--GDP_GER(-1)	0.996295	0.010610	93.90234	0.0000
_ITA--GDP_ITA(-1)	1.006682	0.008819	114.1450	0.0000
_JPN--GDP_JPN(-1)	1.019567	0.008179	124.6552	0.0000
_UK--GDP_UK(-1)	1.003127	0.015082	66.51012	0.0000
_US--GDP_US(-1)	0.996048	0.018346	54.29327	0.0000
Fixed Effects				
_CAN--C	261.2705			
_FRA--C	266.0358			
_GER--C	318.7881			
_ITA--C	194.7066			
_JPN--C	192.1386			
_UK--C	151.0268			
_US--C	273.2745			

The coefficients on GDP? have lengthy labels formed by taking the cross-section specific variable (for example GDP\_CAN(-1)), and prepending the cross-section identifier to denote the equation for which this coefficient is relevant. For example, the first coefficient is the coefficient of GDP\_CAN(-1) in the \_CAN equation. Note that this prepending is necessary for identification purposes since you could, for example, have included the GDP\_UK(-1) variable as an explanatory variable in the \_JPN equation.

Lastly, we can estimate a model with random effects:



Dependent Variable: GDP?  
Method: GLS (Variance Components)  
Date: 10/20/97 Time: 16:08  
Sample: 1951 1992  
Included observations: 42  
Total panel observations 294

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	237.1665	43.99291	5.391016	0.0000
GDP?(-1)	1.001317	0.004233	236.5521	0.0000
Random Effects				
_CAN--C	-2.714267			
_FRA--C	-2.512245			
_GER--C	10.92250			
_ITA--C	-0.644577			
_JPN--C	27.25335			
_UK--C	-21.98726			
_US--C	-10.31750			
GLS Transformed Regression				
R-squared	0.995206	Mean dependent var	9681.432	
Adjusted R-squared	0.995190	S.D. dependent var	3786.743	
S.E. of regression	262.6320	Sum squared resid	20140863	
Durbin-Watson stat	1.503691			
Unweighted Statistics including Random Effects				
R-squared	0.995238	Mean dependent var	9681.432	
Adjusted R-squared	0.995222	S.D. dependent var	3786.743	
S.E. of regression	261.7485	Sum squared resid	20005585	
Durbin-Watson stat	1.513859			

We have simplified the specification so that GDP?(-1) has a common coefficient. Note that EViews provides estimates of the realization of the random effect components, as well as summary statistics for both the transformed and the untransformed data (where the latter is evaluated after removing the estimated random component). The details on these computations are presented in the “Technical Discussion”, beginning on page 545.

## Pool Equation Views and Procedures

Once you have estimated your pool equation, you can examine your output in the usual ways:

### Representation

Select **View/Representations** to examine your specification. EViews estimates your pool as a system of equations, one for each cross-section unit.

### Estimation Output

View/Estimation Output will change the display to show the results from the pooled estimation.

As with other estimation objects, you can examine the estimates of the coefficient covariance matrix by selecting **View/Coef Covariance Matrix**.

### Testing

EViews allows you to perform coefficient tests on the estimated parameters of your pool equation. Select, **View/Wald Coefficient Tests...** and enter the restriction to be tested.

### Residuals

You can view your residuals in spreadsheet or graphical format by selecting **View/Residuals/Table** or **View/Residuals/Graph**. EViews will display the residuals for each cross-sectional equation. Each residual will be named using the base name RES, followed by the cross-section identifier.

If you wish to save the residuals in series using these names, select **Procs/Make Resids**.

### Residual Covariance/Correlation

You can examine the estimated residual contemporaneous covariance and correlation matrices. Select **View/Residual** and then either **Correlation Matrix** or **Covariance Matrix** to examine the appropriate matrix.

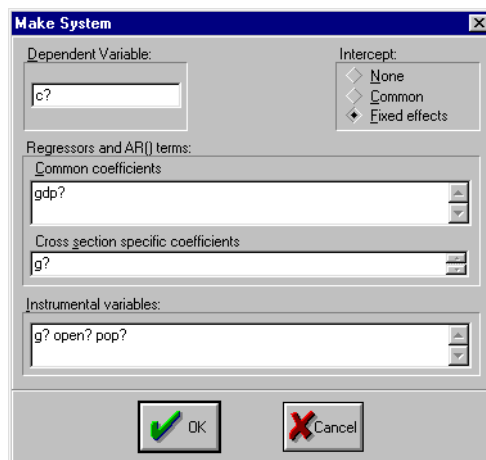
### Forecasting

To perform forecasts using a pool equation you will first make a model. Select **Procs/Make Model** to create an untitled model object that incorporates all of the estimated coefficients. If desired, this model can be edited. Solving the model will generate forecasts for the dependent variable for each of the cross-section units. For further details, see Chapter 23.

### Creating Systems Using Pools

You may have a complex pooled equation specification that cannot be estimated using the pool object. To apply more general estimation methods, such as two-stage least squares, three-stage least squares, and GMM, or to impose arbitrary coefficient restrictions, you should use the pool object to create a system object. You can create a system from an estimated pool, or you can provide information to generate a system from a pool. The system object can be further customized, and estimated using advanced techniques.

Select **Procs/Make System...** and fill out the dialog. You can enter the dependent variable, specify the intercept, common and cross-section specific variables as before. You can also enter a list of instrumental variables. Ordinary variables will appear as an instrument in each equation; pool variables will enter cross-section specific variables as instruments in the corresponding cross-section equation:



## Commands

To create a new pool object, follow the `pool` command keyword with a name for the pool object:

```
pool g7
```

creates a new pool object named G7.

To define the cross-section identifiers of the pool, use the `define` command:

```
g7.define us uk ita ger fra can jap
```

defines the cross section members of the pool object G7 as US, UK, ITA, GER, FRA, CAN, and JAP.

To estimate a pool with fixed effects, use the `f` option:

```
g7.ls(f) cs? @ gdp?
```

estimates a fixed effects model without constraining the coefficients on GDP to be the same for each pool member (which is the same as running OLS for each member separately).

See the *Command and Programming Reference* for a complete list of commands and options available for pool objects.

## Technical Discussion

The class of models that can be estimated using a pool object can be written as

$$y_{it} = \alpha_{it} + x_{it}'\beta_i + \epsilon_{it}, \quad (22.2)$$

where  $y_{it}$  is the dependent variable, and  $x_{it}$  and  $\beta_i$  are  $k$ -vectors of non-constant regressors and parameters for  $i = 1, 2, \dots, N$  cross-sectional units. Each cross-section unit is observed for dated periods  $t = 1, 2, \dots, T$ .

While most of our discussion will be in terms of a balanced sample, EViews does not require that your data be balanced; missing values may be used to represent observations that are not available for analysis in a given period. We will method the behavior of the estimator where there are missing values only where there is the potential for ambiguity.

We can view these data as a set of cross-section specific regressions so that we have  $N$  cross-sectional equations:

$$y_i = \alpha_i + x_i'\beta_i + \epsilon_i \quad (22.3)$$

each with  $T$  observations, stacked on top of one another. For purposes of discussion we will refer to the stacked representation:

$$Y = \alpha + X\beta + \epsilon \quad (22.4)$$

where  $\alpha$ ,  $\beta$ , and  $X$  are set up to include any restrictions on the parameters between cross-sectional units.

The residual covariance matrix for this set of equations is given by:

$$\Omega = E(\epsilon\epsilon') = E \begin{pmatrix} \epsilon_1\epsilon_1' & \epsilon_2\epsilon_1' & \dots & \epsilon_N\epsilon_1' \\ \epsilon_2\epsilon_1' & \epsilon_2\epsilon_2' & & \vdots \\ & & \ddots & \\ \epsilon_N\epsilon_1' & \dots & & \epsilon_N\epsilon_N' \end{pmatrix} \quad (22.5)$$

The basic specification treats the pool specification as a system of equations and estimates the model using system OLS. This specification is appropriate when the residuals are contemporaneously uncorrelated, and time-period and cross-section homoskedastic:

$$\Omega = \sigma^2 I_N \otimes I_T. \quad (22.6)$$

The coefficients and their covariances are estimated using the usual OLS techniques applied to the stacked model.

## Fixed Effects

The fixed effects estimator allows  $\alpha_i$  to differ across cross-section units by estimating different constants for each cross-section. EViews computes the fixed effects by subtracting the “within” mean from each variable and estimating OLS using the transformed data:

$$y_i - \bar{y}_i = (\bar{x}_i - \bar{x}_i')\beta + (\epsilon_i - \bar{\epsilon}_i) \quad (22.7)$$

where  $\bar{y}_i = \sum_t y_{it}/N$ ,  $\bar{x}_i = \sum_t x_{it}/N$ , and  $\bar{\epsilon}_i = \sum_t \epsilon_{it}/N$ .

The coefficient covariance matrix estimates are given by the usual OLS covariance formula applied to the mean differenced model:

$$\text{var}(b_{FE}) = \hat{\sigma}_W^2 (\tilde{X}'\tilde{X})^{-1} \quad (22.8)$$

where  $\tilde{X}$  represents the mean differenced  $X$ , and

$$\hat{\sigma}_W^2 = \frac{e_{FE}' e_{FE}}{NT - N - K} = \frac{\sum_{it} (\tilde{y}_{it} - \tilde{x}_{it}' b_{FE})^2}{NT - N - K}, \quad (22.9)$$

where  $e_{FE}' e_{FE}$  is the SSR from the fixed effects model. If the pool is unbalanced,  $NT$  is replaced by the total number of observations excluding missing values.

The fixed effects themselves are not estimated directly. We report estimated fixed effects computed from

$$\hat{\alpha}_i = \sum_t (\bar{y}_i - \bar{x}_i' b_{FE}) / N \quad (22.10)$$

Standard errors are not reported for the fixed effects coefficients. If you wish to obtain standard errors for the fixed effects, you should re-estimate a model with no intercept, including the constant term as a cross-section specific regressor.

You should note that estimating the cross-section specific constant regression model with a large number of cross-section units may be time-consuming, and may result in estimates that are less accurate than those obtained using the fixed-effects option.

## Random Effects

The random effects model assumes that the term  $\alpha_{it}$  is the sum of a common constant  $\alpha$  and a time-invariant cross-section specific random variable  $u_i$  that is uncorrelated with the residual  $\epsilon_{it}$ . EViews estimates the random effects model using the following steps:

- (1) Use the residuals  $e_{FE}$  from the fixed effects model to estimate the variance of  $\epsilon_{it}$  using  $\hat{\sigma}_W^2$  as described above.
- (2) Estimate the between-group (cross-sectional mean) model and compute:

$$\hat{\sigma}_B^2 = \frac{e_B' e_B}{N - K}, \quad \hat{\sigma}_u^2 = \hat{\sigma}_B^2 - \frac{\hat{\sigma}_W^2}{T} \quad (22.11)$$

where  $e_B' e_B$  is the SSR from the between-group regression. If the estimated  $\hat{\sigma}_u^2$  is negative, EViews will return an error message.

In the case where there are missing observations so that  $T_i$  varies among cross-sections, EViews will use the largest  $T_i$  in computing the variance estimates. This procedure is consistent, provided that the number of missing observations is asymptotically negligible.

(3) Apply OLS to the GLS transformed variables ( $X$  includes the constant term and the regressors  $x$ )

$$y_{it}^* = y_{it} - \hat{\lambda} \bar{y}_i, \quad X_{it}^* = X_{it} - \hat{\lambda} \bar{X}_i \quad (22.12)$$

where  $\hat{\lambda} = 1 - \hat{\sigma}_W / \hat{\sigma}_B$ .

The EViews output displays the parameter estimates for the  $\beta$  derived from (3). The standard errors are computed using the standard estimator of the covariance matrix.

EViews also presents estimates of the values of the random effects. These values are computed using the formula:

$$\hat{u}_i = \frac{\hat{\sigma}_u^2}{\hat{\sigma}_B^2} \left( \sum_t (y_{it} - X_{it} b_{RE}) \right), \quad (22.13)$$

yielding the best linear unbiased predictor of  $u_i$ .

(4) Lastly, EViews displays weighted and unweighted summary statistics. The weighted statistics are from the GLS equation estimated in step (3). The unweighted statistics are derived using residuals from the original model based upon the parameters and the estimated random effect from step (3):

$$\epsilon_{it} = y_{it} - X_{it} b_{RE} - \hat{u}_i. \quad (22.14)$$

## Cross-Section Weighting

Cross-section weighted regression is appropriate when the residuals are cross-section heteroskedastic and contemporaneously uncorrelated:

$$\Omega = E(\epsilon\epsilon') = E \begin{pmatrix} \sigma_1^2 I_{T_1} & 0 & \dots & 0 \\ 0 & \sigma_2^2 I_{T_2} & & \vdots \\ & & \ddots & \\ 0 & \dots & & \sigma_N^2 I_{T_N} \end{pmatrix} \quad (22.15)$$

EViews performs FGLS with  $\hat{\sigma}_i^2$  estimated from a first-stage pooled OLS regression. The estimated variances are: computed as:

$$\hat{\sigma}_i^2 = \sum_{t=1}^{T_i} (y_{it} - \hat{y}_{it})^2 / T_i, \quad (22.16)$$

where  $\hat{y}_{it}$  are the OLS fitted values.

The estimated coefficients values and covariance matrix are given by the standard GLS estimator.

### SUR Weighting

SUR weighted least squares (sometimes referred to as the Parks estimator) is the feasible GLS estimator when the residuals are both cross-section heteroskedastic and contemporaneously correlated (to simplify notation, we assume that  $T_i = T$  for all  $i$ ):

$$\Omega = E(\epsilon\epsilon)' = \begin{pmatrix} \sigma_{11}I_T & \sigma_{12}I_T & \cdots & \sigma_{1N}I_T \\ \sigma_{21}I_T & \sigma_{22}I_T & & \vdots \\ & & \ddots & \\ \sigma_{N1}I_T & \cdots & & \sigma_{NN}I_T \end{pmatrix} = \Sigma \otimes I_T, \quad (22.17)$$

where  $\Sigma$  is the symmetric matrix of contemporaneous correlations:

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1N} \\ \sigma_{21} & \sigma_{22} & & \\ & & \ddots & \\ \sigma_{N1} & \cdots & \cdots & \sigma_{NN} \end{pmatrix}, \quad (22.18)$$

with typical element  $\sigma_{ij} = E(\epsilon_{jt}\epsilon_{it})$ , which is assumed constant across  $t$ .

EViews estimates the SUR model using  $\hat{\sigma}_{ij}$  estimated from a first-stage pooled OLS regression:

$$\hat{\sigma}_{ij} = \sum_t (y_{it} - \hat{y}_{it})^2 / \max(T_i, T_j). \quad (22.19)$$

The use of the max function in the denominator handles the case of unbalanced data by down-weighting the covariance terms. Provided that the number of missing values is asymptotically negligible, this approach yields a consistent estimator of  $\Sigma$  that is generally invertible.

The parameter estimates and the covariance matrix of the parameters of the model are computed using the standard GLS formulae.

### White Covariance Estimation

White's heteroskedasticity consistent covariance estimates may be computed for pooled specifications (except for SUR and random effects estimation). In computing the estimates, EViews calculates the White covariance matrix using the stacked model:

$$\text{var}(b) = \frac{(\sum N_i)}{(\sum N_i) - K} (X'X)^{-1} \left( \sum_{i,t} u_{it}^2 (x_{it}x_{it}') \right) (X'X)^{-1} \quad (22.20)$$

where  $K$  is the total number of estimated parameters. This variance estimator is robust to heteroskedasticity within each cross-section, but does not account for the possibility of contemporaneous correlation across cross-sections.



## Chapter 23. Models

---

A model in EViews is a set of simultaneous equations that are used for forecasting and simulation. Unlike other objects in EViews that contain systems of equations, models do not contain unknown coefficients to be estimated. Instead, models allow you to solve for unknown values for the endogenous variables.

In this chapter we describe various methods of creating model objects in EViews and techniques for using these objects to forecast and perform simulation.

### Creating a Model

To create a model, click **Object/New Object/Model** in the main toolbar. In the dialog that appears, type in a name for your model. When you click **OK**, an empty model window should appear. In the sections below, we will describe how to take this model and add equations and statements to the specification.

There is an alternative approach that deserves mention. Instead of creating an empty model and entering the results of your estimation procedures, you can create a model directly from objects that contain estimated equations. Select **Procs/Make Model** from the equation, VAR, system, pool, or state space toolbars or menus, and EViews will create a new model containing the estimated equations. You may then enter additional equations or statements to this model.

### Specifying a Model

Models in EViews consist of two components: equations and statements.

#### Entering Equations

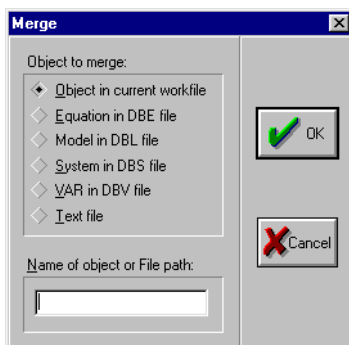
There are four ways to enter equations in a model:

- Enter equations by typing or cutting-and-pasting directly into the model window. We discuss the details of model equation specification below.

Note that you can always see the behavioral equations associated with an EViews estimation object by examining the representations view. Open the object, select **View/Representations**, and examine the listing under **Substituted Coefficients**. The substituted coefficients representation can be cut-and-pasted into your model file.

- Merge objects containing estimated equations from the workfile or disk. Click on the **Merge** button in the model toolbar and specify the object name to merge. You

can merge equations from equation, VAR, system, pool, or even from other model objects. To merge equations from objects containing estimated equations, push the **Merge** button on the model window's toolbar and the following dialog appears:



If the object is in the current workfile, select the top button and type the name of the equation, system, VAR, pool, or model, in the edit box. If the object is in an EViews databank file, or if you want to merge in text from a text file, select the appropriate button and type the file name, including an optional path specification, in the field box. When you click **OK**, the information from the object or file will be copied into the model.

- Make links to objects containing equations. To link an equation into your model, simply enter a colon followed by the name of the equation, VAR, system, pool, or state space object:

:cseq

If the equations in your model have been estimated using an EViews object, you are strongly encouraged to link equations, rather than entering or cutting and pasting the actual specification. The advantage of linking is that when you re-estimate the linked object, the equation in the model will be updated automatically.

- You can create a model directly from objects containing estimated equations. From the object window, select **Procs/Make Model**. EViews will place the results from the estimated equation in a new model object. You may then add to the model using one of the previous methods of entering equations in a model.

## Equations in a Model

The equations in a model can be behavioral equations or identities. Any parameter values must be specified as part of the behavioral equation. For example,

$$y = 0.02207 + 0.58048*y(-1) + 0.33048*y(-2)$$

is a valid behavioral equation, while,

$$y = \text{eq1.c}(1) + \text{eq1.c}(2)*y(-1) + \text{eq1.c}(3)*y(-2)$$

uses the elements in the coefficient vector associated with equation object EQ1.

Behavioral equations may be nonlinear in both variables and coefficients, and may contain ARMA components. More generally, any equation that can be estimated in EViews may appear in a model.

## Nonlinear Specifications and Expressions

When you are building your model, you are not restricted to linear models. For example, the following is a valid equation for a model:

$$s\_k = (1 + .2 * \log(p\_k) + .4 * \log(p\_1)) / (1 - .1 * \log(p\_k))$$

Of course, some nonlinearities will create models that have no solutions or models that cannot be solved using standard computational methods.

Normally, equations have a single series name to the left of the equal sign. This series is the endogenous variable determined by the equation. But you can also put simple formulas to the left of the equal sign. EViews will take the first series name (or potential series name) in the formula as the endogenous variable determined by that equation. This means that you can use EViews to solve for implicit equations in models. For example, if the equation is

$$\log(x) = z$$

the series X will be calculated by solving  $X = \exp(Z)$ . The following are all legal equations that solve for X:

$$\begin{aligned} 1/x &= 2.54 + 3.12*y \\ \log(x/y)/14.4 &= 2.782 + 3.0*z \\ 2+x+3*z &= 4*time \\ d(x) &= nrnd \end{aligned}$$

There are some equations with formulas to the left of the equal sign that EViews cannot solve. See Chapter 5 for more details on the use of expressions in EViews.

## ARMA specifications

When you enter an equation with an ARMA error term, you will provide the values of the AR, MA, and SMA coefficients. If you merge an equation with ARMA errors, EViews will take care of the coefficient values. However, if you wish to type in the equation yourself, you should specify the coefficient values in square brackets (without spaces) at the end of the equation. For MA and SMA terms, you can also include a backcast date. The date indicates how far back to go in computing the moving average innovations. When you solve a model with MA backcasting, you will need to supply data all the way back to the backcast date.

Here are two examples of equations with ARMA error terms:

```
unemp=2.45-2.564*dmr+[ar(1)=.5432,ar(2)=.2134]
dev=+[ar(1)=.65,ma(1)=.32,backcast=1950]
```

Take note of the leading “+” sign in the second equation. You must include this term if the ARMA terms are at the beginning of the equation.

If you specify ARMA coefficients that are nonstationary or noninvertible, the model may not be solvable.

### Add factors

If you link an equation by a colon, you can also specify an add factor. Put a + or - after the name of the equation and a number or the name of a series to be added or subtracted from the equation. For example, if you have estimated an equation of the form

```
equation eq1.ls y c x
```

where the results are stored in an equation object named EQ1, then adding the line

```
:eq1 - 500
```

to the model will include the equation in the model with the estimated coefficients, with the intercept term reduced by 500.

Note that if the equation contains an expression rather than a simple series as the dependent variable, then the add factor will be applied to the equation *after* the equation has been normalized by EViews into its explicit form. That is, if the equation named EQ2 has the representation:

```
log(x) = .3423 + 4.3423 * x
```

then including the equation with the line:

```
:eq2 + 2
```

will solve the model using the transformed equation:

```
x = exp(.3423 + 4.3423 * x) + 2
```

### Identities

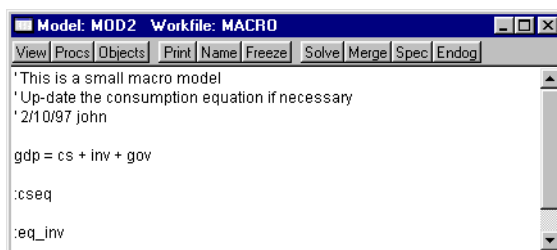
In addition to the behavioral equations, some models may have identities that allow the model to be solved for the endogenous variables. Before trying to solve the model, make sure that there are exactly as many equations in the model as there are endogenous variables to solve for. You must set up your equations so that each equation has a different endogenous variable as the first series on the left-hand side. Sometimes you may need to rewrite identities or behavioral equations in order to put your model in this form.

## Statements in a Model

In addition to equations, you can include statements in models. These statements can be used to control various aspects of the model solution procedure.

### Comments

To include comments in the model, precede the comment with a single quote mark ('). Everything to the right of the single quote will be ignored by EViews. For example,



The first three lines, each preceded by a single quote, are comments that are ignored by EViews.

### Assign Statement

In normal operation, the solution procedure puts the model solutions into the series that appear as endogenous variables in the model. If you solve over a span of observations that includes historical data, the historical data will be lost as they are overwritten by the solved values.

To avoid destroying your historical data, you should include an `assign` statement to store the solved series under a different name. Include a line with the `assign` command, followed by the name of the endogenous variable and a corresponding name to hold the solved series. For example, the statement

```
assign gdp gdpf
```

instructs EViews to put the solved values of GDP into a series named GDPF. You may include as many pairs of series names as you want in a single `assign` statement. For example,

```
assign gdp gdpf cs csf inv invf
```

Two special keywords in an `assign` statement allow you to assign names to *all* of the solved variables without having to list each endogenous variable explicitly. `assign` followed by the word `@all` and one or several letters, instructs EViews to append the listed letters to the original endogenous variable name and assign it as the name for the solved series. For example, if you include the statement

```
assign @all f
```

the solution for GNP will be put into the series GNPF, the solution for CPI will be put into CPIF, and so on. If you include

```
assign @all _v1
```

the solution for GNP will be put into the series GNP\_V1, and the solution for CPI will be placed in CPI\_V1, etc.

Similarly, `assign` followed by the keyword `@prefix` and a list of characters means to prefix the original endogenous variable name with the list of characters, and to use this as the name for the solved series. For example, if you include the statement

```
assign @prefix frc_
```

the solution for GNP will be put into the series FRC\_GNP.

Only one global `@all` or `@prefix` `assign` statement may appear in a model. However, global assignment of names may be overridden by explicit assignment of names for individual variables. For example, the statements

```
assign @all f gnp gnp_fr
```

or

```
assign gnp gnp_fr @all f
```

or

```
assign gnp gnp_fr  
assign @all f
```

all append the letter F to the original variable name and assign it as the name of all solved series, with the exception of GNP, whose solved series will be named GNP\_FR.

To make the syntax of the `assign` statement compatible with the other model control statements, all of which begin with an `@` sign (see below), you may optionally use the word `@assign` in place of `assign`. For example,

```
@assign gnp frgnp  
@assign @all f
```

If you are low on memory, you may not be able to use `assign` statements for all of the series in a large model. If so, be sure to protect your original data by saving your workfile or storing the series to disk.

By default, all models create by EViews from existing objects will have the

```
assign @all f
```

statement placed at the beginning of the model file.

## Exclude Statement

The exclude statement specifies equations to be excluded from the model solution. Follow the word `@exclude` by a list of endogenous variables to be excluded. For example, the statement:

```
@exclude gnp unemp
```

causes the equations with GNP and UNEMP on the left-hand side to be excluded from model solution. GNP and UNEMP are treated as exogenous variables for purposes of simulation.

`@exclude` can also drop model equations over selected samples. This exclusion has the effect of making EViews treat the exogenous variables as exogenous for the relevant period. Include the observations to be excluded as sample range pairs between the word `@exclude` and the list of endogenous variables. For example,

```
@exclude 73.1 78.4 gnp unemp
```

excludes the equations for GNP and UNEMP (treats GNP and UNEMP as exogenous) when simulating the period 1973:1–1978.4 and includes the equations for GNP and UNEMP for all other periods.

`@exclude` is most useful for the following situation. Suppose data for some of the endogenous variables in the solution period are available and you want to use the actual data for those variables in solving for the other endogenous variables. Then simply exclude those variables from the solution only for the period that data are available within the solution period.

You may add as many exclude statements as you need in a single model. Each of these exclude statements may specify different sample ranges.

## Trace Statement

The trace statement allows you to monitor the solution process. EViews will display a table containing the intermediate solutions for each variable requested, providing iteration specific information that may aid you in solving convergence problems.

To trace the solution process, enter the keyword `@trace` by a list of endogenous variable names that you wish to monitor. For example,

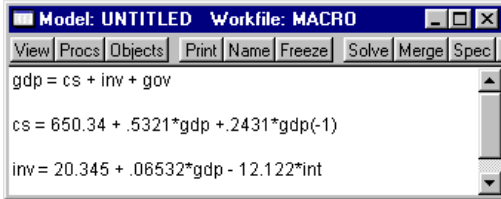
```
@trace gdp cpi
```

displays a table of the solved values of GDP and CPI at each iteration.

If you are having convergence difficulties, note that EViews provides a number of options for controlling the model solution iteration algorithm. These options are described in greater detail below.

## Example

Here is an example of a simple macro model with three equations.

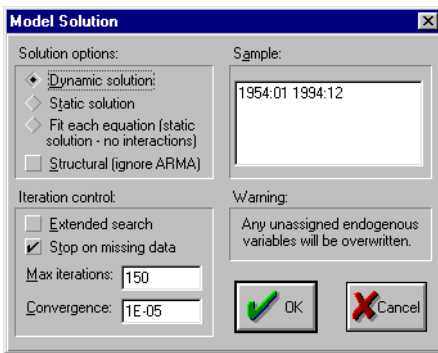


The three variables GDP, CS, INV on the left-hand side of each equation are the endogenous variables in this model. There should be one equation for each endogenous variable in the model. The exogenous variables that appear only on the right-hand side are INT and GOV. The first equation is an identity and the second two are behavioral equations, presum-

ably estimated by two-stage least squares. When you solve the model, EViews solves for the endogenous variables, given data for the exogenous variables.

## Solving Models

Once you have specified the equations and statements in the model, you are ready to solve the model. Before you solve the model, make sure that there are as many equations as there are endogenous variables to be solved. To solve the model, simply click the **Solve** button in the model toolbar. You should see the following dialog box offering various options for controlling the solution process.



We will describe these options shortly; most of the time you can leave them at their default values. Make sure to set the sample to the period you wish to solve the model; the default is the current workfile sample. Note also the warning message; if you do not include an assign statement in the model, all endogenous series will be replaced by the solved series.

When the model is solved, EViews will save the solved values in the workfile as ordinary series. These series may be worked with like any other series in EViews.

## Solution Methods

**Solve** offers the following three solution options:

- **Dynamic solution** performs true *multi-step* forecasts. If you select this option, EViews uses historical data for lagged endogenous variables if they are dated prior to the first period of the simulation. Thereafter it uses the values forecasted by the model itself.



- **Static solution** provides a series of *one-period ahead* forecasts. In static solution, EViews solves for all current endogenous variables, using actual values for all lagged variables in the model.
- **Fit each equation** is a variant of static solution. Fitting each equation separately solves for the dependent variable in each equation using actual values for all current and lagged variables in the equation. Since actual values, rather than solved values, are used for the right-hand side endogenous variables, EViews ignores current interactions between the variables in the model.

This method gives the same results as a single equation static forecast on each equation of the model.

If there are no ARMA terms, EViews will assume that the errors of the model are zero when solving for the endogenous variables.

If the model contains ARMA terms, EViews evaluates the lagged errors in your equations using data prior to the solution period. Then EViews will use the ARMA error structure in solving the model, after setting the innovations to zero. See Chapter 14, “Forecasting with ARMA errors”, page 320, for a detailed discussion of EViews handling of ARMA errors in forecasting.

For some types of work, you may wish to assume that the errors in your equations are always zero, even if there are ARMA terms. The **Structural (ignore ARMA)** option solves the model assuming the errors are zero. The **Structural** option can be used with any of the three solution methods described above.

## Computational Algorithm

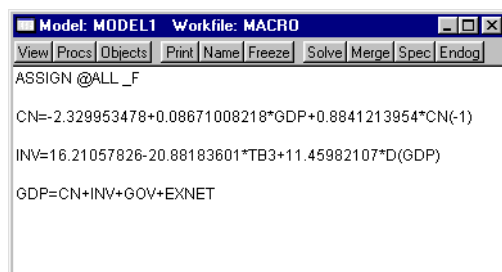
EViews solves both linear and nonlinear models using an iterative Gauss-Seidel method. Gauss-Seidel works by evaluating each equation in the order that it appears in the model, and uses the new value of the left-hand variable in an equation as the value of that variable when it appears in any later equation.

The algorithm is therefore dependent on the order of the equations in your model. This has two important implications for model solution:

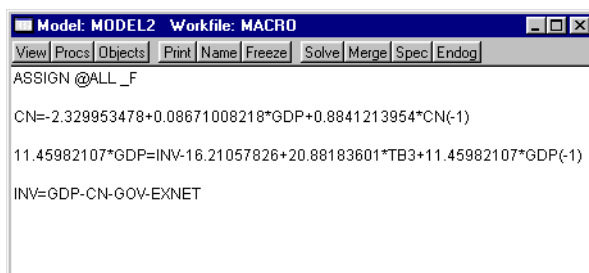
- You must set up your equations so that each equation has a different endogenous variable on the left-hand side. Sometimes you may need to rewrite identities or behavioral equations in order to put your model in this form.
- If EViews has problems solving your model, you should try to rewrite its equations so that the most important variable in each equation is on the left-hand side. In solving your model, EViews will reorder the equations so that ones that have no right-hand endogenous variables come first. In this way, when later equations are evaluated, the correct values for at least some of the right-hand variables will be used from the start. By the same token, you

should put equations early in the model if their right-hand endogenous variables are relatively unimportant.

For example, an attempt to solve the model



returned an overflow message for the INV equation, presumably because of the large coefficients. However, a solution may be obtained once we use the identity and rewrite the model as:



## Iteration control

For most models, the Gauss-Seidel procedure will converge rapidly and provide you with a solution without difficulty. However, the procedure is not foolproof. There are some models for which Gauss-Seidel cannot find the solution. Some types of nonlinearities and other characteristics will result in models that cannot be solved.

## Extended Search

If one or more coefficients of the right-hand variables are very large in economic terms, the solution may not converge. If you select **Extended search** option in the Model Solution dialog box, the procedure can sometimes find the solution to a model it might otherwise be incapable of finding. This option causes the procedure to use smaller step sizes, moving towards its solution in smaller increments at each iteration. While this procedure is slower than the EViews default method, it is slightly better at finding solutions to difficult models.

## Maximum Iterations

For some models, the procedure will make progress, but will not reach the solution within the pre-specified number of iterations. The default is a maximum of 150 iterations. You may use the **Max iterations** option to increase the number of iterations.

## Convergence Criterion

You may use the **Convergence** option to change the convergence criterion. Convergence occurs when the percent change in each solved variable during an iteration is less than the criterion. The default value for the criterion is .00001 or .001 percent. You may specify a smaller value if you require more accurate results or a larger value if you want the procedure to run faster and are willing to accept lower accuracy. Unfortunately, there is no strict relation between the percent change in a variable from one iteration to the next and the accuracy of the solution.

## Working with Models

### Using a model as a multivariate series generator

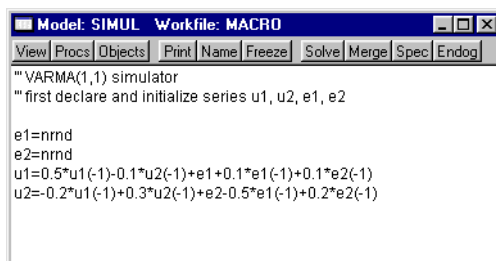
You can use models as a multivariate series generator. For example, suppose for purposes of simulation, that you want to generate the following bivariate ARMA process:

$$\begin{bmatrix} u_{1,t} \\ u_{2,t} \end{bmatrix} = \begin{bmatrix} 0.5 & -0.1 \\ -0.2 & 0.3 \end{bmatrix} \begin{bmatrix} u_{1,t-1} \\ u_{2,t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix} + \begin{bmatrix} 0.1 & 0.1 \\ -0.5 & 0.2 \end{bmatrix} \begin{bmatrix} \epsilon_{1,t-1} \\ \epsilon_{2,t-1} \end{bmatrix} \quad (23.1)$$

where  $\epsilon_1$ ,  $\epsilon_2$  are independent standard normal variables. Note that you cannot generate  $u_1$  and  $u_2$  separately using **Quick/Generate Series...**, since each series depends on the lagged value of the other series. To generate  $u_1$  and  $u_2$ , first declare and set the initial values of the innovations and the two series by the commands

```
series e1=0
series e2=0
series u1=0
series u2=0
```

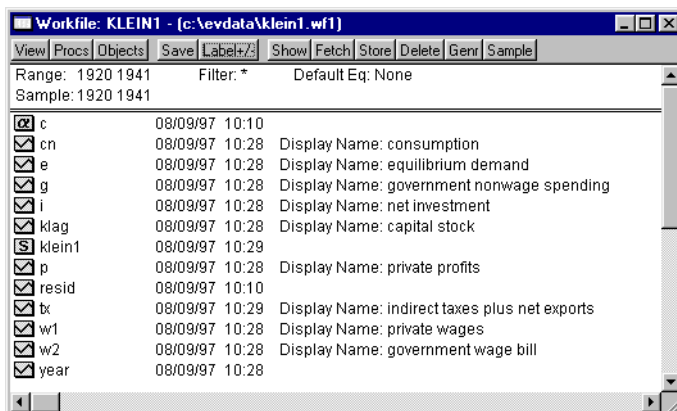
Then create and specify a model as follows:



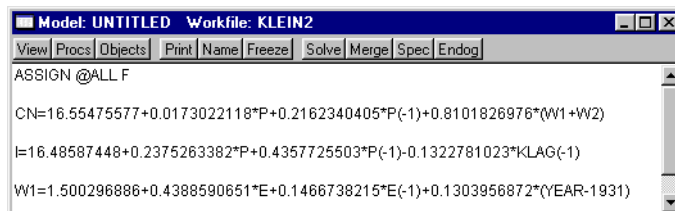
Note that we do not include an ASSIGN statement: since we are generating new series we want the solved values to be placed in the endogenous series. Now simply solve the model to generate the desired series.

## Forecasting from Models

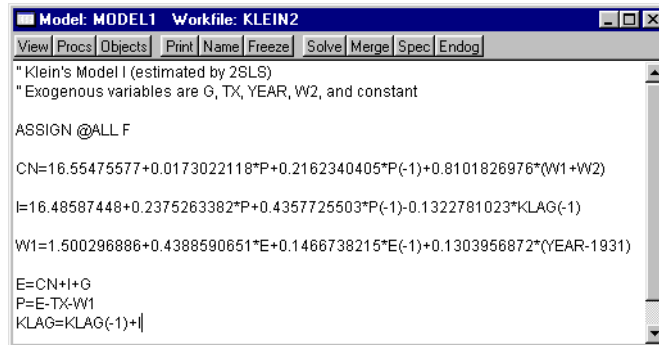
As an illustration, we simulate the well-known Klein's Model I under various scenarios. The series names used in the model are:



We first estimate the model by TSLS in a system object. Choosing **Procs/Make Model** from the system toolbar produces the following model:



Note that EViews includes an assignment statement for models created as a procedure from other estimated objects. Klein's Model I consists of these three behavioral equations and three other identities. We type in the identities directly into the model window as follows:



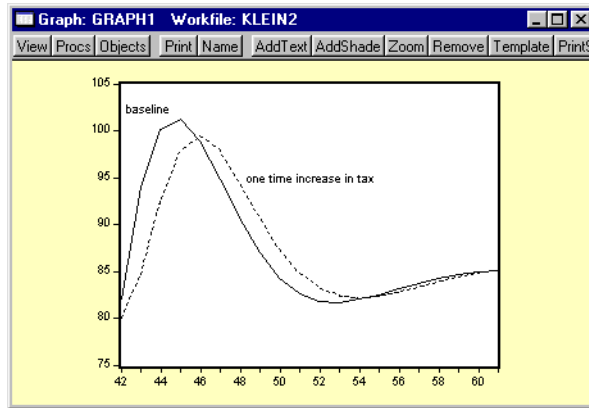
The model was estimated using Klein's original annual data from 1920 to 1941. We simulate the model to obtain forecasts for the period 1942–1961. To do this, first expand the workfile range by choosing **Procs/Expand Range...** from the main toolbar. To solve the model, we need forecasts of the exogenous variables G, TX, YEAR, W2 over the solution period. Although one can fit univariate ARMA models or apply exponential smoothing to obtain forecasts, we simply fix the values of the exogenous variables (except the linear trend YEAR) at the end of sample value in 1941. The following commands generate the values of the exogenous variables for the forecasting period:

```
smpl 42 61
series g=g(-1)
series tx=tx(-1)
series w2=w2(-1)
series year=@trend(1942)+1942
```

To solve the model, simply click **Solve** in the model toolbar. You can repeat the simulation under various scenarios and compare with the baseline forecasts. For example, to see what happens if there were a one time increase in taxes ,TX, at the beginning of the forecast sample, modify the value of TX by the commands

```
smpl 42 42
series tx=tx+10
series 42 61
```

then resolve the model. If you want to compare the results with the previous simulation, do not forget to modify the `assign` statement to save the simulated series under a different name; otherwise, the previous results will be overwritten. The forecasts of consumption under the two scenarios are given by:



The tax increase reduces consumption in the first four years and then overshoots the baseline level for about eight years. The effect of the one time increase in tax disappears in about twelve years.

## Commands

To create a new model object, follow the `model` command with a name for the model:

```
model simul1
```

creates and opens the window of a new model named `SIMUL1`.

To solve a model, use the `solve` procedure:

```
simul1.solve
```

See the *Command and Programming Reference* for a complete list of commands and options available for model objects.

## Appendix A. Mathematical Operators and Functions

---

### Operators

All of the operators described below may be used in expressions involving series and scalar values. When applied to a series expression, the operation is performed for each observation in the current sample. The precedence of evaluation is listed on page 82 of the *User's Guide*. Note that you can enforce order-of-evaluation using appropriate parentheses

Expression	Operator	Description
+	add	$x+y$ adds the contents of X and Y
-	subtract	$x-y$ subtracts the contents of Y from X
*	multiply	$x*y$ multiplies the contents of X by Y
/	divide	$x/y$ divides the contents of X by Y
^	raise to the power	$x^y$ raises X to the power of Y
>	greater than	$x>y$ takes the value 1 if X exceeds Y, and 0 otherwise
<	less than	$x<y$ takes the value 1 if Y exceeds X, and 0 otherwise
=	equal to	$x=y$ takes the value 1 if X and Y are equal, and 0 otherwise
<>	not equal to	$x<>y$ takes the value 1 if X and Y are not equal, and 0 if they are equal
<=	less than or equal to	$x<=y$ takes the value 1 if X does not exceed Y, and 0 otherwise
>=	greater than or equal to	$x>=y$ takes the value 1 if Y does not exceed X, and 0 otherwise
and	logical and	$x$ and $y$ takes the value 1 if both X and Y are nonzero, and 0 otherwise
or	logical or	$x$ or $y$ takes the value 1 if either X or Y is nonzero, and 0 otherwise

### Date and Observation Functions

These functions allow you to identify the period associated with a given observation, or the value associated with a given date/observation.

Function	Name	Description
@day	observation day	for daily or weekly workfiles, returns the observation day in the month for each observation
@elem(x, "d") , @elem(x, s)	element	returns the value of the series <i>x</i> , at observation or date, <i>d</i> , or string <i>s</i> .
@month	observation month	returns the month of observation (for monthly, daily, and weekly data) for each observation
@quarter	observation quarter	returns the quarter of observation (except for annual, semi-annual, and undated data) for each observation
@year	observation year	returns the year associated with each observation (except for undated data) for each observation

For example, if you create the series

```
series y = @month
```

with a monthly workfile, Y will contain a numeric indicator for each month (1 through 12). If you create the series

```
series z = @quarter
```

for the same workfile, EViews will fill Z with the numeric quarter indicator (1 through 4) associated with each observation.

You may only use these functions in a workfile of at least as high frequency. Thus, while you can use the @month function in a daily or annual workfile, you cannot use the @day function in a monthly or annual workfile, nor can you use the @month function in an annual workfile.

## Basic Mathematical Functions

These functions perform basic mathematical operations. When applied to a series, they return a value for every observation in the current sample. When applied to a matrix object, they return a value for every element of the matrix object. The functions will return NA values for observations where the input values are NAs, or where the input values are not valid. For example, the square-root function @sqrt, will return NAs for all observations less than zero.

Note that the logarithmic functions are base-*e* (natural logarithms). To convert the natural logarithm into  $\log_{10}$ , use the relationship:  $\log_{10}(x) = \log_e(x)/\log_e(10) \approx \log_e(x)/2.303$ .



Name	Function	Examples/Description
@abs(x), abs(x)	absolute value	@abs(-3) = 3
@ceiling(x)	smallest integer not less than	@ceiling(2.34) = 3; @ceiling(4) = 4
@exp(x), exp(x)	exponential, $e^x$	@exp(1) $\cong$ 2.71813
@fact(x)	factorial, $x!$	@fact(3) = 6; @fact(0) = 1
@factlog(x)	natural logarithm of the factorial, $\log_e(x!)$	@factlog(3) $\cong$ 1.7918; @factlog(0) = 0
@floor(x)	largest integer not greater than	@floor(1.23) = 1; @floor(3) = 3
@inv(x)	reciprocal, $1/x$	@inv(2) = 0.5
@mod(x, y)	floating point remainder	returns the remainder of x/y with the same sign as x. If y = 0, the result is 0.
@log(x), log(x)	natural logarithm, $\log_e(x)$	@log(2) $\cong$ 0.693 ; log(2.71813) $\cong$ 1
@nan(x, y)	recode NAs in X to Y	returns x if x < > na, and y if x = na
@recode(s, x, y)	recode by condition	returns x if condition s is true; otherwise returns y: @recode(y > 0, x, 2)
@round(x)	round to the nearest integer	@round(-97.5) = -98; @round(3.5) = 4
@sqrt(x), sqr(x)	square root	@sqrt(9) = 3

## Time Series Functions

The following functions facilitate working with time series data. Note that NAs will be returned for observations for which lagged values are not available. For example, `d(x)` returns a missing value for the first observation in the workfile, since the lagged value is not available.

Name	Function	Description
d(x)	first difference	$(1 - L)X = X - X(-1)$ , where $L$ is the lag operator
d(x, n)	$n$ -th order difference	$(1 - L)^n X$
d(x, n, s)	$n$ -th order difference with a seasonal difference at $s$	$(1 - L)^n (1 - L^s) X$
dlog(x)	first difference of the logarithm	$(1 - L)\log(X)$ $= \log(X) - \log(X(-1))$
dlog(x, n)	$n$ -th order difference of the logarithm	$(1 - L)^n \log(X)$

<code>dlog(x, n, s)</code>	$n$ -th order difference of the logarithm with a seasonal difference at $s$	$(1 - L)^n (1 - L^s) \log(X)$
<code>@movav(x, n)</code>	$n$ -period backward moving average	$\text{@movav}(x, 3) = (X + X(-1) + X(-2))/3$
<code>@movsum(x, n)</code>	$n$ -period backward moving sum	$\text{@movsum}(x, 3) = (X + X(-1) + X(-2))$
<code>@pch(x)</code>	one-period percentage change (in decimal)	$(X - X(-1))/(X(-1))$
<code>@pcha(x)</code>	one-period percentage change—annualized (in decimal)	$\text{@pcha}(x) = (1 + \text{@pch}(x))^n - 1$ where $n$ is the lag associated with one-year ( $n = 4$ for quarterly data, etc.)
<code>@pchy(x)</code>	one-year percentage change (in decimal)	$(X - X(-n))/(X(-n))$ , where $n$ is the lag associated with one-year ( $n = 12$ for annual data, etc.)
<code>@seas(n)</code>	seasonal dummy	returns 1 when the quarter or month equals $n$ and 0 otherwise
<code>@trend,</code> <code>@trend(n)</code>	time trend	returns a trend series, normalized to 0 in period $n$ , where $n$ is a date or observation number; if $n$ is omitted, then the series is normalized at the first observation in the workfile

## Descriptive Statistics

These functions compute descriptive statistics for the current sample, excluding missing values. To determine the number of observations available for a given series, use the `@obs` function. Note that where appropriate, EViews will perform casewise exclusion of data with missing values. For example, `@cov(x, y)` and `@cor(x, y)` will use only observations for which data on both  $X$  and  $Y$  are valid.

These functions compute descriptive statistics for the specified sample, excluding missing values if necessary. The default sample is the current workfile sample. When computing the values using a series, you can specify a sample as the last argument of the descriptive statistic function, either as a string (in double quotes) or using the name of a sample object. For example, “series  $z = \text{@mean}(x, "1945:01\ 1979:12")$ ” or “ $w = \text{@var}(y, s2)$ ” where  $s2$  is the name of a sample object.

To determine the number of observations available for a given series, use the `@obs` function. Note that where appropriate, EViews will perform casewise exclusion of data with missing values. For example, `@cov(x,y)` and `@cor(x,y)` will use only those observations for which data on both X and Y are available.

In the following table, arguments in square brackets [ ] are optional arguments:

- [s]: sample expression in double quotes or name of a sample object. For `@quantile`, you must provide the method option argument in order to include a sample argument.

Function	Name	Description
<code>@cor(x,y[,s])</code>	correlation	the correlation between X and Y
<code>@cov(x,y[,s])</code>	covariance	the covariance between X and Y
<code>@inner(x,y[,s])</code>	inner product	the inner product of X and Y
<code>@obs(x[,s])</code>	number of observations	the number of non-missing observations for X in the current sample
<code>@mean(x[,s])</code>	mean	average of the values in X
<code>@median(x[,s])</code>	median	computes the median of the X (uses the average of middle two observations if the number of observations is even)
<code>@min(x[,s])</code>	minimum	minimum of the values in X
<code>@max(x[,s])</code>	maximum	maximum of the values in X
<code>@quantile(x,q[,m,s])</code>	quantile	the q-th quantile of the series X. m is an optional argument that specifies the following methods (see CDF-options - 1 (rankit - default), 2 (ordinary), 3 (van der Waerden), 4 (Blom), 5 (Tukey)
<code>@stdev(x[,s])</code>	standard deviation	square root of the unbiased sample variance (sum-of-squared residuals divided by $n - 1$ )
<code>@sum(x[,s])</code>	sum	the sum of X
<code>@sumsq(x[,s])</code>	sum-of-squares	sum of the squares of X
<code>@var(x[,s])</code>	variance	variance of the values in X (division by $n$ )

## Additional and Special Functions

EViews provides a number of utility and “special” functions. A number of these functions are used in evaluating the properties of various statistical distributions or for returning special mathematical values such as Euler’s constant. For further details on special functions, see the extensive discussions in Temme (1996), Abramowitz and Stegun (1964), and Press, *et al.* (1992).

Function	Description
@beta(a,b)	beta integral (Euler integral of the second kind) $B(a,b) = \int_0^1 t^{a-1}(1-t)^{b-1} dt = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ for $a, b > 0$ .
@betainc(x,a,b)	incomplete beta integral $\frac{1}{B(a,b)} \int_0^x t^{a-1}(1-t)^{b-1} dt$ for $0 \leq x \leq 1$ , and $a, b > 0$ .
@betaincder(x,a,b,s)	derivative of the incomplete beta integral: Evaluates the derivatives of the incomplete beta integral $B(x,a,b)$ , where $s$ is an integer from 1 to 9 corresponding to the desired derivative: $\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} = \begin{bmatrix} \frac{\partial B}{\partial x} & \frac{\partial B}{\partial a} & \frac{\partial B}{\partial b} \\ \frac{\partial^2 B}{\partial x^2} & \frac{\partial^2 B}{\partial x \partial a} & \frac{\partial^2 B}{\partial x \partial b} \\ \frac{\partial^2 B}{\partial a^2} & \frac{\partial^2 B}{\partial a \partial b} & \frac{\partial^2 B}{\partial b^2} \end{bmatrix}$
@betainv(p,a,b)	inverse of the incomplete beta integral: for $0 \leq p \leq 1$ , and $a, b > 0$ , returns an $x$ satisfying: $p = \frac{1}{B(a,b)} \int_0^x t^{a-1}(1-t)^{b-1} dt$
@betalog(a,b)	natural logarithm of the beta integral $\log B(a,b) = \log \Gamma(a) + \log \Gamma(b) - \log \Gamma(a+b)$
@binom(n,x)	binomial coefficient $\binom{n}{x} = \frac{n!}{x!(n-x)!}$
@binomlog(n,x)	natural logarithm of the binomial coefficient $\log(n!) - \log(x!) - \log((n-x)!)$
@cloglog(x)	complementary log-log function $\log(-\log(1-x))$ See also @qextreme.

@digamma(x), @psi(x)	first derivative of the log gamma function
$\psi(x) = \frac{d \log \Gamma(x)}{dx} = \frac{1}{\Gamma(x)} \frac{d\Gamma(x)}{dx}$	
@erf(x)	error function
$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt, \text{ for } x \geq 0.$	
@erfc(x)	complementary error function
$\operatorname{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-t^2} dt = 1 - \operatorname{erf}(x), \text{ for } x \geq 0.$	
@gamma(x)	(complete) gamma function
$\Gamma(x) = \int_0^\infty e^{-t} t^{x-1} dt, \text{ for } x \geq 0$	
@gammader(x)	first derivative of the gamma function:
$\Gamma'(x) = d\Gamma(x)/(dx)$	
Note: Euler's constant $\gamma \cong 0.5772$ may be evaluated by $\gamma = -\Gamma'(1)$ . See also @digamma and @trigamma.	
@gammainc(x, a)	incomplete gamma function
$G(x, a) = \frac{1}{\Gamma(a)} \int_0^x e^{-t} t^{a-1} dt$	
for $x \geq 0$ and $a > 0$ .	
@gammaincder(x, a, n)	derivative of the incomplete gamma function:
Evaluates the derivatives of the incomplete gamma integral $G(x, a)$ , where $s$ is an integer from 1 to 5 corresponding to the desired derivative:	
$\begin{bmatrix} 1 & 2 & - \\ 3 & 4 & 5 \end{bmatrix} = \begin{bmatrix} \frac{\partial G}{\partial x} & \frac{\partial G}{\partial a} & - \\ \frac{\partial^2 G}{\partial x^2} & \frac{\partial^2 G}{\partial x \partial a} & \frac{\partial^2 G}{\partial a^2} \end{bmatrix}$	

@gammaincinv(p, a)	inverse of the incomplete gamma function: find the value of $x$ satisfying $p = G(x, a) = \frac{1}{\Gamma(a)} \int_0^x e^{-t} t^{a-1} dt$ for $0 \leq p \leq 1$ and $a > 0$ ,
@gammalog(x)	logarithm of the gamma function: $\log \Gamma(x)$ For derivatives of this function see @digamma and @trigamma.
@logit(x)	logistic transform $\frac{1}{1 + e^{-x}} = \frac{e^x}{1 + e^x}$
@psi(x)	see @digamma
@trigamma(x)	second derivative of the log gamma function $\psi'(x) = \frac{d^2 \log \Gamma(x)}{dx^2}$

## Trigonometric Functions

When applied to a series, all of the trigonometric functions operate on every observation in the current sample and return a value for every observation. Where relevant, the input and results should/will be expressed in radians. All results are real valued—complex values will return NAs.

Function	Name	Examples/Description
@acos(x)	arc cosine (real results in radians)	@ acos(−1) ≅ π
@asin(x)	arc sine (real results in radians)	@ asin(1) = π/2
@atan(x)	arc tangent (results in radians)	@ atan(1) = π/4
@cos(x)	cosine (argument in radians)	@ cos(3.14159) ≅ −1
@sin(x)	sine (argument in radians)	@ sin(3.14159) ≅ 0
@tan(x)	tangent (argument in radians)	@ tan(1) ≅ 1.5574

## Statistical Distribution Functions

The following functions provide access to the density or probability functions, cumulative distribution, quantile functions, and random number generators for a number of standard statistical distributions.

There are four functions associated with each distribution. The first character of each function name identifies the type of function:

Function Type	Beginning of Name
Cumulative distribution (CDF)	@c
Density or probability	@d
Quantile (inverse CDF)	@q
Random number generator	@r

The remainder of the function name identifies the distribution. For example, the functions for the beta distribution are @cbeta, @dbeta, @qbeta and @rbeta.

When used with series arguments, EViews will evaluate the function for each observation in the current sample. As with other functions, NA or invalid inputs will yield NA values. For values outside of the support, the functions will return zero.

Note that the CDFs are assumed to be right-continuous:  $F_X(k) = \Pr(X \leq k)$ . The quantile functions will return the smallest value where the CDF evaluated at the value equals or exceeds the probability of interest:  $q_X(p) = q^*$ , where  $F_X(q^*) \geq p$ . The inequalities are only relevant for discrete distributions.

The information provided below should be sufficient to identify the meaning of the parameters for each distribution. For further details, see the *Command and Programming Reference*.

Distribution	Functions	Density/Probability Function
Beta	@cbeta(x, a, b)	$f(x, a, b) = \frac{x^{a-1}(1-x)^{b-1}}{B(a, b)}$
	@dbeta(x, a, b)	
	@qbeta(p, a, b)	for $0 \leq p \leq 1$ and $a, b > 0$ , where $B$ is the @beta function defined above.
	@rbeta(a, b)	
Binomial	@cbinom(x, n, p)	$\Pr(x, n, p) = \binom{n}{x} p^x (1-p)^{n-x}$
	@dbinom(x, n, p)	
	@qbinom(s, n, p)	if $x = 0, 1, \dots, n$ , and 0 otherwise, for $0 \leq p \leq 1$ .
	@rbinom(n, p)	
Chi-square	@cchisq(x, v)	$f(x, v) = \frac{1}{2^{v/2} \Gamma(v/2)} x^{v/2-1} e^{-x/2}$
	@dchisq(x, v)	
	@qchisq(p, v)	where $x \geq 0$ , and $v > 0$ . Note that the degrees-of-freedom parameter $v$ need not be an integer.
	@rchisq(v)	

Exponential	<code>@cexp (x, m)</code> <code>@dexp (x, m)</code> <code>@qexp (p, m)</code> <code>@rexp (m)</code>	$f(x, m) = \frac{1}{m} e^{-x/m}$ <p>for <math>x \geq 0</math> , and <math>m &gt; 0</math> .</p>
Extreme Value (Type I-minimum)	<code>@cextreme (x)</code> <code>@dextreme (x)</code> <code>@qextreme (p)</code> , <code>@cloglog (p)</code> <code>@rextreme</code>	$f(x) = \exp(x - e^x)$ <p>for <math>-\infty &lt; x &lt; \infty</math> .</p>
F-distribution	<code>@cdfdist (x, v1, v2)</code> <code>@dfdistr (x, v1, v2)</code> <code>@qfdistr (p, v1, v2)</code> <code>@rfdistr (v1, v1)</code>	$f(x, v_1, v_2) = \frac{v_1^{v_1/2} v_2^{v_2/2}}{B(v_1/2, v_2/2)}$ $x^{(v_1-2)/2} (v_2 + v_1 x)^{-(v_1 + v_2)/2}$ <p>where <math>x \geq 0</math> , and <math>v_1, v_2 &gt; 0</math> . Note that the functions allow for fractional degrees-of-freedom parameters <math>v_1</math> and <math>v_2</math> .</p>
Gamma	<code>@cgamma (x, b, r)</code> <code>@dgamma (x, b, r)</code> <code>@qgamma (p, b, r)</code> <code>@rgamma (b, r)</code>	$f(x, b, r) = b^{-r} x^{r-1} e^{-x/b} / \Gamma(r)$ <p>where <math>x &gt; 0</math> , and <math>b, r &gt; 0</math> .</p>
Laplace	<code>@claplace (x)</code> <code>@dlaplace (x)</code> <code>@qlaplace (x)</code> <code>@rlaplace</code>	$f(x) = \frac{1}{2} e^{- x }$ <p>for <math>-\infty &lt; x &lt; \infty</math> .</p>
Logistic	<code>@clogistic (x)</code> <code>@dlogistic (x)</code> <code>@qlogistic (p)</code> <code>@rlogistic</code>	$f(x) = \frac{1}{1 + e^{-x}} = \frac{e^x}{1 + e^x}$ <p>for <math>-\infty &lt; x &lt; \infty</math> .</p>
Log-normal	<code>@clognorm (x, m, s)</code> <code>@dlognorm (x, m, s)</code> <code>@qlognorm (p, m, s)</code> <code>@rlognorm (m, s)</code>	$f(x, m, s) = \frac{1}{x \sqrt{2\pi s^2}} e^{-(\log x - m)^2 / (2s^2)}$ <p><math>x &gt; 0</math> , <math>-\infty &lt; m &lt; \infty</math> , and <math>s &gt; 0</math> .</p>



Negative Binomial	<code>@cnegbin(x,n,p)</code> <code>@dnegbin(x,n,p)</code> <code>@qnegbin(s,n,p)</code> <code>@rnegbin(n,p)</code>	$\Pr(x, n, p) = \frac{\Gamma(x+n)}{\Gamma(x+1)\Gamma(n)} p^n (1-p)^x$ <p>if <math>x = 0, 1, \dots, n, \dots</math>, and 0 otherwise, for <math>0 \leq p \leq 1</math>.</p>
Normal (Gaussian)	<code>@cnorm(x)</code> <code>@dnorm(x)</code> <code>@qnorm(p)</code> <code>@rnorm, rnd</code>	$f(x) = (2\pi)^{-1/2} e^{-x^2/2}$ <p>for <math>-\infty &lt; x &lt; \infty</math>.</p>
Poisson	<code>@cpoisson(x,m)</code> <code>@dpoisson(x,m)</code> <code>@qpoisson(p,m)</code> <code>@rpoisson(m)</code>	$\Pr(x, m) = m^x e^{-m} / x!$ <p>if <math>x = 0, 1, \dots, n, \dots</math>, and 0 otherwise, for <math>m &gt; 0</math>.</p>
Pareto	<code>@cpareto(x,a,k)</code> <code>@dpareto(x,a,k)</code> <code>@qpareto(p,a,k)</code> <code>@rpareto(a,k)</code>	$f(x, a, k) = (ak^a) / x^{a+1}$ <p>for <math>a &gt; 0</math>, and <math>x \geq k \geq 0</math>.</p>
Student's <i>t</i> -distribution	<code>@ctdist(x,v)</code> <code>@dtdist(x,v)</code> <code>@qtdist(p,v)</code> <code>@rtdist(v)</code>	$f(x, v) = \frac{\Gamma((v+1)/2)}{(v\pi)^{1/2} \Gamma(v/2)} (1 + (x^2/v))^{-\frac{(v+1)}{2}}$ <p>for <math>-\infty &lt; x &lt; \infty</math>, and <math>v &gt; 0</math>. Note that <math>v = 1</math> yields the Cauchy distribution.</p>
Uniform	<code>@cunif(x,a,b)</code> <code>@dunif(x,a,b)</code> <code>@qunif(p,a,b)</code> <code>@runif(a,b), rnd</code>	$f(x) = \frac{1}{b-a}$ <p>for <math>a \leq x \leq b</math>, and <math>b &gt; a</math></p>
Weibull	<code>@cweib(x,m,a)</code> <code>@dweib(x,m,a)</code> <code>@qweib(p,m,a)</code> <code>@rweib(m,a)</code>	$f(x, m, a) = am^{-a} x^{a-1} e^{-(x/m)^a}$ <p>where <math>-\infty &lt; x &lt; \infty</math>, and <math>m, a &gt; 0</math>.</p>

## Additional Distribution Functions

The following utility functions were designed to facilitate the computation of *p*-values for common statistical tests. While these results may be derived using the distributional functions above, they are retained for convenience and backward compatibility.

Function	Distribution	Description
@chisq(x, v)	Chi-square	Returns the probability that a Chi-squared statistic with $d$ degrees of freedom exceeds $x$ :  $1 - @cchisq(x, d)$
@fdist(x, v1, v2)	$F$ -distribution	Probability that an $F$ -statistic with $v_1$ numerator degrees of freedom and $v_2$ denominator degrees of freedom exceeds $X$ :  $1 - @cfdist(x, v1, v2)$
@tdist(x, v)	$t$ -distribution	Probability that a $t$ -statistic with $v$ degrees of freedom exceeds $x$ in absolute value (two-sided $p$ -value):  $2 * (1 - @ctdist(x, v))$

## Single Equation Regression Statistics

There are a number of functions that return scalar values for equation results, either for the default, or a named equation. These scalar values may be used in generating new series.

The functions are listed here for completeness: there is a full discussion of the use of these statistics in Chapter 11.

Function	Description
@aic	Akaike information criterion: $-2(l/T) + 2(k/T)$
@coefs(i)	coefficient $i$
@cov(i, j)	covariance of coefficients $i$ and $j$
@dw	Durbin-Watson statistic
@f	$F$ -statistic
@jstat	scalar containing the value of the objective function (for GMM)
@logl	value of the log-likelihood function $l$
@meandep	mean of the dependent variable
@ncoef	total number of estimated coefficients
@r2	R-squared statistic
@rbar2	adjusted R-squared statistic
@regobs	number of observations in regression
@schwarz	Schwarz information criterion: $-2(l/T) + k \log(T)/T$
@sddep	standard deviation of the dependent variable
@se	standard error of the regression

@ssr	sum of squared residuals
@stderrs(i)	standard error for coefficient i
@tstats(i)	t-statistic value for coefficient i

## Systems of Equations Regression Statistics

Name	Description
@coefs(i)	coefficient i (except for VARs)
@cov(i, j)	covariance of coefficients i and j
@dw(j)	Durbin-Watson statistic for equation j
@jstat	scalar containing the value of the objective function (for GMM)
@logl	value of the log-likelihood function (where appropriate)
@meandep(j)	mean of the dependent variable for equation j
@ncoef(j)	number of coefficients in equation j
@ncross	number of cross-sections in the pool (Pools only)
@ncrossest	number of cross-sections used in last pool estimation (Pool only)
@neqn	Number of equations in system
@r2(j)	R-squared statistic for equation j
@rbar2(j)	adjusted R-squared statistic for equation j
@regobs(j)	number of observations in regression for equation j
@residcov(i, j)	covariance of residuals from equation i with those in equation j in a VAR or system object. @residcov must be preceded with a named object, e.g. var1.@residcov(2, 2)
@sddep(j)	standard deviation of the dependent variable for equation j
@se(j)	standard error of the regression for equation j
@ssr(j)	sum of squared residuals for equation j
@stderrs(i)	standard error for coefficient i (except for VARs)
@sysncoef	total number of coefficients in the system
@tstats(i)	t-statistic value for coefficient i (except for VARs)

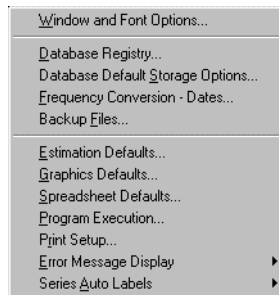


## Appendix B. Global Options

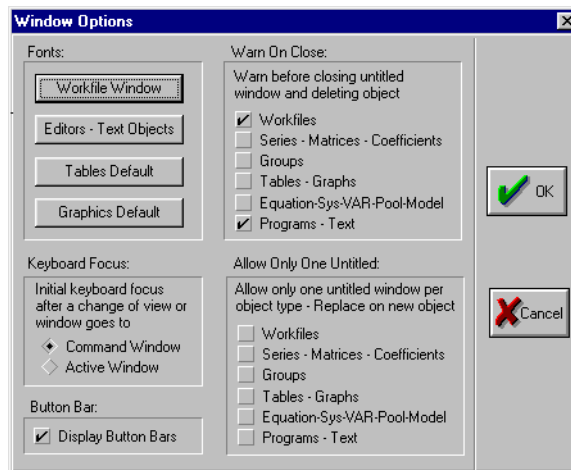
---

### Setting Global Options

The **Options** menu in the main toolbar controls the default behavior of many of the operations in EViews.



### Window and Font Options



- The **Fonts** options allow you to change the default font styles and sizes for each object type. Press the button of the type of object for which you want to change the font and select the new font from the Font dialog.
- **Keyboard Focus** controls where the keyboard cursor is placed when you change views or windows. The two choices make a difference only when the window is an editable window such as the specification view of system objects.

- The **Button Bar** toggles on and off the display of the button bar in the object windows. The button bar provides shortcuts to frequently used items on the main menu. The main menu is always displayed, so that if you turn off the button bar, you can still perform all operations from the main menu.
- **Warn On Close** instructs EViews to provide a warning message when you close an untitled object. You may choose to set warnings for various object types. By default, EViews warns you that closing an unnamed object without naming it will cause the object to be deleted. Along with the warning, you will be given an opportunity to name the object. If you turn the warning off, closing an untitled window will automatically delete the object.
- **Allow Only One Untitled** specifies, for each object type, whether to allow multiple untitled objects to be opened at the same time. If only one is allowed, creating a new untitled object will cause any existing untitled object of that type to be deleted automatically.

Setting EViews to allow only one untitled object reduces the number of windows that you will see on your desktop. By default, only one untitled workfile, group, equation, system, model, program or text object can be open, but any number of untitled series, matrices, coefficients, tables, or graphs may be open.

If you allow only one untitled object, we strongly recommend that you select **Warn on Close**. With this option set, you will be given the opportunity to name and save an existing untitled object, otherwise the object will simply be deleted.

## Backup Files

You can choose whether to keep backup copies of workfiles, data bank files, and program files. The backup copy will have the same name as the file, but with the first character in the extension changed to ~. For example, if you have a workfile named MYDATA.WF1, the backup file will be named MYDATA.~F1.

## Error Message Display

You can choose to have error messages displayed prominently in a box (the default), or less conspicuously, on the status line at the bottom of the EViews window.

## Series Auto Labels

By default, EViews keeps a history in the series label of the commands that created or modified a series. You can choose to turn this option off.

## Estimation Defaults

You can set the maximum number of iterations and convergence criterion for iterative estimation methods globally. These settings will be used as default in iterative estimation methods for equation, log likelihood, state space, and system objects. Note that previously estimated EViews

objects that were estimated with the previous default settings will be unchanged, unless reestimated.

## Graphics Default

These options control how the graphs appear when you first create a graph. The options are explained in Chapter 10, “Graphs, Tables, and Text Objects”.

## Spreadsheet Defaults

These options control the default spreadsheet view of series and group objects. This is the view that is displayed when the object is created.

For series spreadsheet views, the default is to display the series for the entire workfile range in a single column. You may change to display the series only for the current sample or to display the series in multiple columns. You also have the option whether to include the series label in the spreadsheet view or not and whether to display the view with edit mode enabled.

For group spreadsheet views, you have the option to display the spreadsheet in transposed form, with series arranged in rows instead of columns. You can also choose to open the view with edit mode enabled by default.

Note that these options only apply to newly created series or group objects. Once you modify the view using the **Edit**, **Smpl**, **Label**, **Wide**, or **Transpose** options, the object will be displayed accordingly.

## Dates-Frequency

For daily and weekly data, the default notation for dates is the American Month/Day/Year. You can switch to the European notation, where the day precedes the month.

You can also set the default frequency conversion method to use when you move data to lower frequency workfiles. The frequency conversion methods are discussed in detail in Chapter 6, “EViews Databases”.

## Program Execution

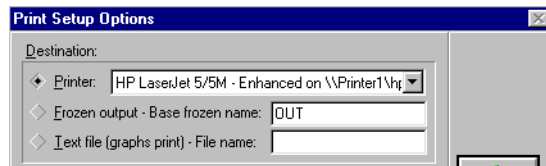
By default, EViews runs programs in verbose mode, listing commands in the status line as they are executed. Alternatively, you can switch to quiet mode, which suppresses this information. EViews will run faster in quiet mode since the status line display does not need to be updated.

The default may always be overridden from the Run Program dialog, or by using the option “Q” in the run statement, as in:

```
run(q) myprogram
```

## Print Setup

These options determine the default print behavior when you push the **Print** button or select **View/Print Selected** from the window toolbar.



The top of the Print Setup Options dialog gives you three choices for the destination of printed output. The default is to send the view or procedure output to the current Windows printer. The other two options redirect output into graph or table objects, or into text files.

## Output Redirection

The **Frozen Output** option saves the views or output from procedures as frozen graphs, tables, or text objects in the workfile. No output is sent to the printer. Under this option, the **Print** button behaves as if it were the **Freeze** button.

If you choose **Frozen Output**, you must also supply a base name. Each print command will then create a frozen object within the active workfile, naming it with the base name followed by a number. For example, if you supply the base name of OUT, the first print command will generate a table or graph named OUT01, the second print command will generate OUT02, and so on. You can select some or all of the frozen objects in the workfile window, open them and copy-and-paste to your word processor, print them (after turning off output redirection), store them, or delete them. See Chapter 1 and Chapter 10 for more on copying output to other Windows applications.

The **Text File** option redirects printing of output to a file. You may use this option to create a log of the results from your EViews session. Note that all print commands of graph views or graph objects are still sent to the printer.

If you choose the **Text File** option, you must supply a file name. The file will be created in the default path with a .TXT extension, unless you explicitly specify the full path name. If a file with the same name already exists, EViews will warn you that the file will be overwritten. Any instruction to print tables or text will append the output to the end of the file.



## Appendix C. Date Formats

---

One of the more important features of EViews is the ability to process dated data. Once you determine the frequency of your data, EViews will use all available calendar information to organize and manage your data.

When using EViews, you must often specify dates. Dates are required when creating a workfile, specifying a sample, or specifying periods in hypothesis tests. This chapter discusses the format for dates in EViews.

### Date Specification

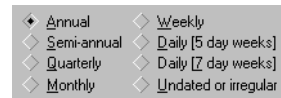
The rules for specifying dates depend on the workfile frequency. Currently, there are eight frequencies available in EViews:

#### Annual

Specify the year. For example, 1981, 1895, 382, or 95.

Years in the 21st and subsequent centuries must be identified with full-year identifiers, e.g. 2020 or 9789 or 50234. Years prior to the 20th century must also be identified with full identifiers, e.g. 382 or 1776 or 1492. Years in the 20th century may be identified using either 2 or 4-digit identifiers, e.g. 97 or 1997. We recommend that you use 4-digit identifiers to avoid confusion.

Note that because 2-digit identifiers are assumed to be in the 20th century, EViews cannot handle dates prior to A.D. 100.



#### Semi-annual

Follow the year number with a colon and the number 1 for the first half of the year and the number 2 for the second half of the year. Examples: 1965:1, 2010:2.

#### Quarterly

Follow the year number with a colon and the quarter number. The quarter number should be 1, 2, 3 or 4. Examples: 1992:1, 1965:4, 2002:3.

#### Monthly

Follow the year number with a colon and the month number. You may prefix the one digit month numbers with a zero, e.g. 01 or 04. Examples: 1956:04, 1965:10, 2010:12.

## Weekly and Daily

By default, the dates are specified in American notation as

`mm/dd/yyyy`

where mm, dd, yyyy stand for the month, date, and year numbers, respectively. The month and date numbers can be a single digit and the year numbers can be less than four digits. Examples: 3/1/1995, 03/01/1995, 10/01/2010, 1/01/1965.

With weekly data, the start of the week is identified by the first day specified in the workfile. All weeks will follow this convention. Thus, if you specify the start of your workfile as 10/15/1997, a Wednesday, all of your weeks will begin on Wednesdays.

In any sample statement, if you specify a date in the middle of a week, EViews will view that date as the week containing the specified date. For example, for the Wednesday-based workfile above, if you enter:

```
smp1 10/17/1997 10/30/1997
```

EViews will display the sample as

```
10/15/1997 10/29/1997
```

since October 17 is the Friday of the week beginning on October 15 and October 30 is the Thursday of the week beginning on October 29.

You can use European notation by choosing **Options/Dates-Frequency...** from the main menu. Dates in European notation have the form

`dd/mm/yyyy`

For example, 8/10/97 in American notation indicates August 10, 1997, while in European notation it indicates October 8, 1997.

## Undated or Irregular

Simply specify the observation number.

## Implicit Dating

When you first create a workfile, you can specify the workfile range by only indicating the year number for any frequency (except undated). EViews will create a workfile that starts at the beginning of the specified year and that ends at the end of the specified year. For example, if you specify the workfile range as

```
60 90
```

EViews will create a workfile with range

1960–1990	for annual
1960:1–1990:2	for semi-annual
1960:1–1990:4	for quarterly
1960:01–1990:12	for monthly
1/01/1960–12/28/1990	for weekly
1/01/1960–12/31/1990	for daily
60–90	for undated or irregular

If you want to create a workfile starting or ending in the middle of a year, you should specify the full date.

## Special Date Functions

There are three special functions that provide shortcuts for changing the workfile sample: `@all`, `@first`, and `@last`.

The keyword `@all` refers to the entire workfile range, while the keywords `@first` and `@last` refer to the first and last observation in the workfile, respectively. For example, suppose you have a workfile with range 1953:01–1996:12. Then to reset the sample to workfile range, you can use any of the following three commands:

```
smp1 @all
smp1 @first 1996:12
smp1 53:1 @last
```

Sample range elements may contain mathematical expressions to create date offsets. This feature can be particularly useful in setting up a fixed width window of observations. For example, to set the sample to the last ten observations of the workfile range, you can use the command

```
smp1 @last-9 @last
```

## Examples

```
1989:2
```

is the second half of 1989 (semi-annual) or the second quarter of 1989 (quarterly) or February 1989 (monthly).

```
1989:02
```

is the same as 1989:2.

1992:4

is the fourth quarter of 1992 (quarterly) or April 1992 (monthly).

82

is the year 1982 (annual) or observation number 82 (undated).

1979:02

is the second half of 1979 (semi-annual) or the second quarter of 1979 (quarterly) or February 1979 (monthly).

1882

is the year 1882 (annual) or observation number 1882 (undated).

1968:5

is May 1968 (monthly).

1976:56

is an error since there is no month 56.

9/30/1996

is September 30, 1986 (daily) or the week starting on that day (weekly).

2/30/1993

is an error since there is no February 30.

## Appendix D. Wildcards

---

EViews supports the use of wildcard characters in a variety of situations where you can enter a list of objects or a list of series. For example, you can use wildcards to:

- specify a group object
- delete a list of objects
- filter the workfile display

This appendix describes some of the issues involved in the use of wildcard characters and expressions.

### Wildcard Expressions

There are two wildcard characters: “\*” and “?”. The wildcard character “\*” matches zero or more characters in a name, and the wildcard “?” matches any single character in a name.

For example, you can use the wildcard expression `GD*` to refer to all objects whose names begin with the characters “GD”. The series `GD`, `GDP`, `GD_F` will be included in this list `GGD`, `GPD` will not. If you use the expression `GD?`, EViews will interpret this as a list of all objects with three character names beginning with the string “GD”: `GDP` and `GD2` will be included, but `GD`, `GD_2` will not.

You can instruct EViews to match a fixed number of characters by using as many “?” wildcard characters as necessary. For example, EViews will interpret `??GDP` as matching all objects with names that begin with any two characters followed by the string “GDP”. `USGDP` and `F_GDP` will be included but `GDP`, `GGDP`, `GDPUS` will not.

You can also mix the different wildcard characters in an expression. For example, you can use the expression `*GDP?` to refer to any object that ends with the string “GDP” and an arbitrary character. Both `GDP_1`, `USGDP_F` will be included.

### Using Wildcard Expressions

Wildcard expressions may be used in filtering the workfile display (see Chapter 3), in selected EViews commands, and in creating a group object.

The following commands support the use of wildcards: `show`, `copy`, `store`, `fetch`, and `delete`.

To create a group using wildcards, simply select **Object/New Objects/Group**, and enter the expression, EViews will first expand the expression, and then attempt to create a group using the corresponding list of series. For example, entering the list

```
y x*
```

will create a group comprised of Y and all series beginning with the letter X. Alternatively, you can enter the command

```
group g1 x* y?? c
```

defines a group G1, consisting of all of the series matching X\*, and all series beginning with the letter Y followed by two arbitrary characters.

If the list contains invalid or incompatible objects, EViews will issue an error message. If X1 is a series and MAT1 is a matrix, the command

```
group g1 x? mat*
```

Once created, these groups may be used anywhere that EViews takes a group as input. For example, if you have a series of dummy variables, DUM1, DUM2, DUM3, ..., DUM9, that you wish to enter in a regression, you can create a group containing the dummy variables, and then enter the group in the regression:

```
group gdum dum?  
equation eq1.ls y x z gdum
```

will run the appropriate regression. Note that we are assuming that the dummy variables are the only series objects which match the wildcard expression DUM?.

## Resolving Ambiguities

Note that an ambiguity can arise with wildcard characters since both “\*” and “?” have multiple uses. The “\*” character may be interpreted as either a multiplication operator or a wildcard character. The “?” character serves as both the single character wildcard and the pool cross section identifier.

### Wildcard versus Multiplication

There is a potential for ambiguity in the use of the wildcard character “\*”.

Suppose you have a workfile with the series X, X2, Y, XYA, XY2. There are then two interpretations of the wildcard expression “X\*2”. The expression may be interpreted as an auto-series representing X multiplied by 2. Alternatively, the expression may be used as a wildcard expression, referring to the series X2 and XY2.

Note that there is only an ambiguity when the character is used in the middle of an expression, not when the wildcard character “\*” is used at the beginning or end of an expression. EViews uses the following rules to determine the interpretation of ambiguous expressions:

- EViews first tests to see whether the expression represents a valid series expression. If so, the expression is treated as an auto-series. If it is not a valid series expression, then EViews will treat the “\*” as a wildcard character. For example,

`Y*X`

`2*X`

are interpreted as auto-series, while

`*X`

`X*A`

are interpreted as wildcard expressions.

- You can force EViews to treat “\*” as a wildcard by preceding the character with another “\*”. Thus, expressions containing “\*\*” will always be treated as wildcard expressions. For example, the expression

`X**2`

unambiguously refers to all objects with names beginning with “X” and ending with “2”. Note that the use of “\*\*” does *not* conflict with the EViews exponentiation operator “^”.

- You can instruct EViews to treat “\*” as a series expression operator by enclosing the expression (or any subexpression) in parentheses. For example,

`(Y*X)`

always refers to X times Y.

We *strongly* encourage you always to resolve the ambiguity by using parentheses to denote series expressions, and double asterisks to denote wildcards (in the middle of expressions), whenever you create a group. This is especially true when group creation occurs in a program; otherwise the behavior of the program will be difficult to predict since it will change as the names of other objects in the workfile change.

## Wildcard versus Pool Identifier

The “?” wildcard character is used both to match any single character in a pattern and as a place-holder for the cross-section identifier in pool objects.

EViews resolves this ambiguity by not allowing the wildcard interpretation of “?” in any expression involving a pool object or entered into a pool dialog: “?” is used exclusively as a cross-section identifier. For example, suppose that you have the pool object POOL1. Then, the expression

`pool1.est y? x? c`

is a regression of the pool variable  $Y?$  on the pool variable  $X?$ , and

```
pool1.delete x?
```

deletes all of the series in the pool series  $X?$ . There is no ambiguity in the interpretation of these expressions since they both involve `POOL1`.

Similarly, when used apart from a pool object, the “?” is interpreted as a wildcard character. Thus,

```
delete x?
```

unambiguously deletes all of the series matching  $X?$ .



## Appendix E. Estimation Algorithms and Options

---

EViews estimates a wide variety of nonlinear models, from nonlinear least squares specifications, to maximum likelihood models, to GMM specifications. These nonlinear estimation problems do not have closed form solutions and must be estimated using iterative procedures. Similarly, EViews solves systems of non-linear equations. Again, there are no closed form solutions to these problems, and EViews must use an iterative method to obtain a solution.

In this appendix, we provide details on the algorithms used by EViews in dealing with these problems.

Our discussion here is necessarily brief. For additional details we direct you to the quite readable discussions in Press, *et al.* (1992), Quandt (1983), Thisted (1988), and Amemiya (1983).

### Nonlinear Optimization Algorithms

The problem faced in non-linear estimation is to find the values of parameters  $\theta$  that optimize (maximize or minimize) an objective function  $F(\theta)$ .

Iterative optimization algorithms work by taking a initial set of values for the parameters, say  $\theta_{(0)}$ , then performing calculations based on these values to obtain a better set of parameter values,  $\theta_{(1)}$ . This process is repeated for  $\theta_{(2)}$ ,  $\theta_{(3)}$  and so on until the objective function  $F$  no longer improves between iterations.

There are three main parts to the optimization process: (1) obtaining the initial parameter values, (2) updating the candidate parameter vector  $\theta$  at each iteration, and (3) determining when we have reached the optimum.

If the objective function is globally concave so that there is a single maximum, any algorithm which improves the parameter vector at each iteration will eventually find this maximum (assuming that the size of the steps taken does not become negligible). If the objective function is not globally concave, different algorithms may find different local maxima, but all iterative algorithms will suffer from the same problem of being unable to tell apart a local and a global maximum.

The main thing that distinguishes different algorithms is how quickly they find the maximum. Unfortunately, there are no hard and fast rules. For some problems, one method may be faster, for other problems it may not. EViews uses a number of different algorithms, and will often let you choose which method you would like to use.

The following sections outline these methods. The algorithms may be broadly classified into three types: *first derivative* methods, *second derivative* methods, and *derivative free* methods. First derivative methods form candidate values using only the current parameter values and the first derivatives of the objective function. Second derivative methods evaluate both the first and the second derivatives of the objective function during the iteration process. As the name suggests, derivative free methods do not compute derivatives.

## Second Derivative Methods

For binary, ordered, censored, and count models, EViews can estimate the model using Newton-Raphson or *quadratic hill-climbing*.

### Newton-Raphson

Candidate values for the parameters  $\theta_{(1)}$  may be obtained using the method of Newton-Raphson by linearizing the first order conditions  $\partial F/\partial \theta$  at the current parameter values,  $\theta_{(i)}$ :

$$\begin{aligned} g_{(i)} + H_{(i)}(\theta_{(i+1)} - \theta_{(i)}) &= 0 \\ \theta_{(i+1)} &= \theta_{(i)} - H_{(i)}^{-1}g_{(i)} \end{aligned} \quad (5.1)$$

where  $g$  is the gradient vector  $\partial F/\partial \theta$ , and  $H$  is the Hessian matrix  $\partial^2 F/\partial \theta^2$ .

If the function is quadratic, Newton-Raphson will find the maximum in a single iteration. If the function is not quadratic, the success of the algorithm will depend on how well a local quadratic approximation captures the shape of the function.

### Quadratic hill-climbing (Goldfeld-Quandt)

This method, which is a straightforward variation on Newton-Raphson, is sometimes attributed to Goldfeld and Quandt. Quadratic hill-climbing modifies the Newton-Raphson algorithm by adding a correction matrix (or ridge factor) to the Hessian. The quadratic hill-climbing updating algorithm is given by

$$\theta_{(i+1)} = \theta_{(i)} - \tilde{H}_{(i)}^{-1}g_{(i)} \quad \text{where } -\tilde{H}_{(i)} = -H_{(i)} + \alpha I \quad (5.2)$$

where  $I$  is the identity matrix and  $\alpha$  is a positive number that is chosen by the algorithm.

The effect of this modification is to push the parameter estimates in the direction of the gradient vector. The idea is that when we are far from the maximum, the local quadratic approximation to the function may be a poor guide to its overall shape, so we may be better off simply following the gradient. The correction may provide better performance at locations far from the optimum, and allows for computation of the direction vector in cases where the Hessian is near singular.

For models which may be estimated using second derivative methods, EViews uses quadratic hill-climbing as its default method. You may elect to use traditional Newton-Raphson, or the first derivative methods described below, by selecting the desired algorithm in the Options menu.

Note that asymptotic standard errors are always computed from the unmodified Hessian once convergence is achieved.

## First Derivative Methods

Second derivative methods may be computationally costly since we need to evaluate the  $k(k + 1)/2$  elements of the second derivative matrix at every iteration. An alternative is to employ methods which require only the first derivatives of the objective function at the parameter values.

For general nonlinear models (nonlinear least squares, ARCH and GARCH, nonlinear system estimators, GMM), EViews uses first derivative methods: Gauss-Newton/BHHH or Marquardt.

### Gauss-Newton/BHHH

This algorithm follows Newton-Raphson, but replaces the negative of the Hessian by an approximation formed from the sum of the outer product of the gradient vectors for each observation's contribution to the objective function. For least squares and log-likelihood functions, this approximation is asymptotically equivalent to the actual Hessian when evaluated at the parameter values which maximize the function. When evaluated away from the maximum, this approximation may be quite poor.

The algorithm is referred to as *Gauss-Newton* for general nonlinear least squares problems, and Berndt, Hall, Hall, and Hausman (*BHHH*) for maximum likelihood problems.

The advantages of approximating the negative Hessian by the outer product of the gradient are that (1) we need to evaluate only the first derivatives, and (2) the outer product is necessarily positive definite. The disadvantage is that, away from the maximum, this approximation may provide a poor guide to the overall shape of the function, so that more iterations may be needed for convergence.

### Marquardt

The Marquardt algorithm modifies the Gauss-Newton algorithm in exactly the same manner as quadratic hill climbing modifies the Newton-Raphson method (by adding a correction matrix (or ridge factor) to the Hessian approximation).

The ridge correction handles numerical problems when the outer product is near singular and may improve the convergence rate. As above, the algorithm pushes the updated parameter values in the direction of the gradient.

For models which may be estimated using first derivative methods, EViews uses Marquardt as its default method. You may elect to use traditional Gauss-Newton via the Options menu.

Note that asymptotic standard errors are always computed from the unmodified (Gauss-Newton) Hessian approximation once convergence is achieved.

## Numerical versus Analytical Derivatives

There are two approaches to evaluating derivatives: numerical and analytical. Where practical, EViews uses hand-calculated analytic derivatives. For highly nonlinear models (such as GARCH), or models where users may specify an arbitrary function (nonlinear least squares, GMM), EViews computes the derivatives numerically using a four-point routine based on Richardson extrapolation.

Analytic derivatives may be faster and more accurate than numeric derivatives, especially if the calculation of analytic derivatives has been carefully optimized to remove common subexpressions. Numeric derivatives will sometimes involve fewer function evaluations, and in some cases may be faster.

## Choosing the step size

At each iteration we can search along the given direction for the optimal step size. EViews performs a simple trial-and-error search at each iteration to determine a step size  $\lambda$  that improves the objective function. This procedure is sometimes referred to as *squeezing* or *stretching*.

Note that  $\lambda$  is not optimized at each iteration since the computation of the direction vector is often more important than the choice of the step size. It is possible, however, that EViews will be unable to find a step size that improves the objective function. In this case, EViews will issue an error message.

EViews also performs a trial-and-error search to determine the scale  $\alpha$  on the correction factor for Marquardt and quadratic hill-climbing methods.

## Derivative free methods

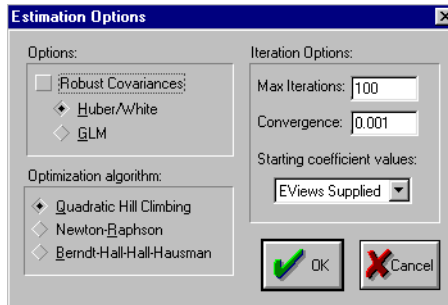
Other optimization routines do not require the computation of derivatives. The *grid search* is a leading example. Grid search simply computes the objective function on a grid of parameter values and chooses the parameters with the highest values. Grid search is computationally costly, especially for multi-parameter models.

EViews uses (a version of) grid search for the exponential smoothing routine.

## Setting Estimation Options

When you estimate an equation in EViews, you will usually enter information into the Equation Specification dialog. The **Options** button in the dialog brings up an Estimation Options dialog that provides you various options to control the estimation procedure. The dialogs will differ depending upon the options available for a particular estimation procedure.

The Options dialog for binary models is shown below:



## Starting Values

Iterative estimation procedures require starting values for the coefficients of the model. There are no general rules for selecting starting values for parameters. Obviously, the closer to the true values, the better, so if you have reasonable guesses for parameter values, these can be useful. In some cases, you can obtain starting values by estimating a restricted version of the model. In general, however, you may have to experiment to find good starting values.



EViews follows three basic rules for selecting starting values:

- For nonlinear least squares type problems, EViews uses the values in the coefficient vector at the time you begin the estimation procedure as starting values.
- For system estimators and ARCH, EViews uses starting values based upon preliminary single equation OLS or TSLs estimation.
- For selected estimation techniques (binary, ordered, count, censored and truncated) EViews has built-in algorithms for determining the starting values using specific information about the objective function. These will be labeled in the Starting Values combo box as **EViews Supplied**.

In the latter two cases, you may change this default behavior by selecting an item from the **Starting coefficient values** drop down menu. You may choose fractions of the default starting values, zero, or arbitrary **User Supplied**.

If you select **User Supplied**, EViews will use the values stored in the C coefficient vector at the time of estimation as starting values. To see the starting values, double click on the coefficient vector in the workfile directory. If the values appear to be reasonable, you can close the window and proceed with estimating your model.

If you wish to change the starting values, first make certain that the spreadsheet view of the coefficient vector is in edit mode, then enter the coefficient values. When you are finished setting the initial values, close the coefficient vector window and estimate your model.

You may also set starting coefficient values from the command window using the **PARAM** command. Simply enter the **PARAM** keyword, followed by pairs of coefficients and their desired values:

```
param c(1) 153 c(2) .68 c(3) .15
```

sets  $C(1) = 153$ ,  $C(2) = .68$ , and  $C(3) = .15$ . All of the other elements of the coefficient vector are left unchanged.

Lastly, if you want to use estimated coefficients from another equation, select **Procs/Update Coefs from Equation** from the equation window toolbar.

For nonlinear least squares problems or situations where you specify the starting values, bear in mind that:

- The objective function must be defined at the starting values. For example, if your objective function contains the expression  $1/C(1)$ , then you cannot set  $C(1)$  to zero. Similarly, if the objective function contains  $\text{LOG}(C(2))$ , then  $C(2)$  must be greater than zero.
- A poor choice of starting values may cause the nonlinear least squares algorithm to fail. EViews begins nonlinear estimation by taking derivatives of the objective function with respect to the parameters, evaluated at these values. If these derivatives are not well behaved, the algorithm may be unable to proceed.

If, for example, the starting values are such that the derivatives are all zero, you will immediately see an error message indicating that EViews has encountered a “Near Singular Matrix”, and the estimation procedure will stop.

- Unless the objective function is globally concave, iterative algorithms may stop at a local optimum. There will generally be no evidence of this fact in any of the output from estimation.

If you are concerned with the possibility of local optima, you may wish to select various starting values and see whether the estimates converge to the same values. One common suggestion is to estimate the model and then randomly alter each of the estimated coefficients by some percentage, then use these new coefficients as starting values in estimation.

### Iteration and Convergence options

There are two common iteration stopping rules: based on the change in the objective function or on the change in parameters. EViews stops iterating when the maximum of the percentage changes in the estimated parameters is smaller than a prespecified level. This is generally a conservative rule since the change in the objective function may be quite small as we approach the optimum (this is how we choose the direction), while the parameters may still be changing.

EViews may stop iterating even when convergence is not achieved. This can happen for two reasons. First, the number of iterations may have reached the prespecified upper bound. In this case,

you should reset the maximum number of iterations to a larger number and try iterating until convergence is achieved.

Second, EViews may issue an error message indicating a “Failure to improve” after certain iterations. This means that even though the parameters change, EViews could not find a direction or step size that improves the objective function. This can happen when the objective function is ill-behaved; make sure your model is identified. You should try other starting values and see if you can approach the optimum from other directions.

Except in rare circumstances, you should not change the convergence criterion. The estimation process achieves convergence if the maximum change in the coefficients (scaled by the corresponding coefficient value) is below the specified value. For example, the default convergence criterion is .001, which means that convergence is achieved if the maximum of the percentage changes in the coefficients is smaller than .1 percent.

## Nonlinear Equation Solution Methods

EViews uses the Gauss-Seidel method when solving systems of nonlinear equations. Suppose the system of equations is given by:

$$\begin{aligned}x_1 &= f_1(x_1, x_2, \dots, x_N, z) \\x_2 &= f_2(x_1, x_2, \dots, x_N, z) \\&\vdots \\x_N &= f_N(x_1, x_2, \dots, x_N, z)\end{aligned}\tag{5.3}$$

where  $x$  are the endogenous variables and  $z$  are the exogenous variables.

The problem is to find a fixed point such that  $x = f(x, z)$ . Gauss-Seidel uses an iterative updating rule of the form:

$$x^{(i+1)} = f(x^{(i)}, z).\tag{5.4}$$

to find the solution. At each iteration, EViews solves the equations in the order that they appear in the model. If an endogenous variable that has already been solved for in that iteration appears later in some other equation, EViews uses the value as solved in that iteration. For example, the  $k$ -th variable in the  $i$ -th iteration is solved by

$$x_k^{(i)} = f_k(x_1^{(i)}, x_2^{(i)}, \dots, x_{k-1}^{(i)}, x_k^{(i-1)}, x_{k+1}^{(i-1)}, \dots, x_N^{(i-1)}, z) \quad .\tag{5.5}$$

Theoretically, the order in which the equations are written should not matter. In practice, the Gauss-Seidel method may fail to converge. In such cases, you should try reordering the equations; list the equations with relatively few and unimportant right-hand side endogenous variables early in the model.

If there are any equations without any right-hand side endogenous variables, EViews evaluates those equations first in any iteration.

## What Does EViews Use?

The following table summarizes the nonlinear algorithms used by EViews:

Type of Estimation Problem	Default method	Optional methods
Nonlinear Least Squares, GARCH, GMM, Nonlinear System, State Space Estimation	Marquardt (numeric)	BHHH (numeric)
Binary, Ordered, Censored, Count Estimation	Quadratic hill-climbing (analytic)	Newton-Raphson (analytic) BHHH (analytic)
Model Solution	Gauss-Seidel	



## Appendix F. Information Criteria

---

As part of the output for most regression procedures, EViews reports various information criteria. The information criteria are often used as a guide in model selection (see for example, Grasa 1989).

The Kullback-Leibler quantity of information contained in a model is the distance from the “true” model and is measured by the log likelihood function. The notion of an information criterion is to provide a measure of information that strikes a balance between this measure of goodness of fit and parsimonious specification of the model. The various information criteria differ in how to strike this balance.

### Definitions

Akaike info criterion (AIC)	$-2(l/T) + 2(k/T)$
Schwarz criterion (SC)	$-2(l/T) + k\log(T)/T$
Hannan-Quinn criterion (HQ)	$-2(l/T) + 2k\log(\log(T))/T$

Consider a regression model with  $k$  parameters, estimated using  $T$  observations. Let  $l$  be the value of the log of the likelihood function using the  $k$  estimated parameters. The various information criteria are all based on  $-2$  times the average log likelihood function, adjusted by a penalty function.

The following information criteria are reported by EViews:

- The Hannan-Quinn criterion is reported only for binary, ordered, censored, and count models.
- The definitions used by EViews may differ slightly from those used by some authors. For example, Grasa (1989, equation 3.21) does not divide the AIC by  $n$ . Other authors omit inessential constants involving  $2\pi$  of the Gaussian log likelihood.

While earlier versions of EViews reported information criteria that omitted inessential constant terms, the current version of EViews always uses the value of the full likelihood function. All of your equation objects estimated in earlier versions of EViews will automatically be updated to reflect this change. You should, however, keep this fact in mind when comparing results from frozen table objects or printed output from previous versions.

- For systems of equations, where applicable, the information criteria are computed using the full system log likelihood. The log-likelihood value is computed assuming a multivariate normal (Gaussian) distribution as:

$$l = -\frac{TM}{2}(1 + \log 2\pi) - \frac{T}{2} \log |\hat{\Omega}| \quad (6.1)$$

where

$$|\hat{\Omega}| = \det\left(\sum_i \hat{\epsilon}\hat{\epsilon}'/T\right) \quad (6.2)$$

$M$  is the number of equations.

## Using Information Criteria as a Guide to Model Selection

As a user of these information criteria as a model selection guide, you select the model with the smallest information criterion.

The information criterion has been widely used in time series analysis to determine the appropriate length of the distributed lag. Lutkepohl (1991, Chapter 4) presents a number of results regarding consistent lag order selection in VAR models.

You should note, however, that the criteria depend on the unit of measurement of the dependent variable  $y$ . For example, you cannot use information criteria to select between a model with dependent variable  $y$  and one with  $\log(y)$ .

## References

---

- Abramowitz, M. and I. A. Stegun (1964) *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, Dover Publications.
- Aitchison, J. and S.D. Silvey (1957) "The Generalization of Probit Analysis to the Case of Multiple Responses," *Biometrika*, 44, 131–140.
- Agresti, Alan (1996) *An Introduction to Categorical Data Analysis*, New York: John Wiley & Sons.
- Andrews, Donald W.K. (1988a) "Chi-Square Diagnostic Tests for Econometric Models: Theory," *Econometrica*, 56, 1419–1453.
- Andrews, Donald W.K. (1988b) "Chi-Square Diagnostic Tests for Econometric Models: Introduction and Applications," *Journal of Econometrics*, 37, 135–156.
- Andrews, Donald W.K. (1991) "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817–858.
- Andrews, Donald W.K. and J. Christopher Monahan (1992) "An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator," *Econometrica*, 60, 953–966.
- Bollerslev, Tim (1986). "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics* 31, 307–327.
- Bollerslev Tim, Ray Y. Chou, and Kenneth F. Kroner (1992) "ARCH Modeling in Finance: A Review of the Theory and Empirical Evidence," *Journal of Econometrics* 52, 5–59.
- Bollerslev, Tim and Jeffrey M. Wooldridge (1992) "Quasi-Maximum Likelihood Estimation and Inference in Dynamic Models with Time Varying Covariances," *Econometric Reviews*, 11, 143–172.
- Bollerslev, Tim, Robert F. Engle and Daniel B. Nelson (1994) "ARCH Models," in Chapter 49 of *Handbook of Econometrics*, Volume 4, North-Holland.
- Bowerman, Bruce L. and Richard T. O'Connell (1979) *Time Series and Forecasting: An Applied Approach*, Duxbury Press.
- Box, George E.P. and Gwilym M. Jenkins (1976) *Time Series Analysis: Forecasting and Control*, Revised Edition, Holden-Day.
- Box, George E.P. and D.A. Pierce (1970) "Distribution of Residual Autocorrelations in Autoregressive Integrated Moving Average Time Series Models," *Journal of the American Statistical Association*, 65, 1509–1526.
- Brown, R.L., J. Durbin, and J.M. Evans (1975) "Techniques for Testing the Constancy of Regression Relationships Over Time," *Journal of the Royal Statistical Society, Series B*, 37, 149–192.
- Brown, M.B. and A.B. Forsythe (1974a) "Robust Tests for the Equality of Variances," *Journal of the American Statistical Association*, 69, 364–367.
- Brown, M.B. and A.B. Forsythe (1974b) "The Small Sample Behavior of Some Test Statistics which Test the Equality of Several Means," *Technometrics*, 16, 129–132.
- Cameron, A. Colin and Pravin K. Trivedi (1990) "Regression-based Tests for Overdispersion in the Poisson Model," *Journal of Econometrics*, 46, 347–364.
- Campbell John Y. and Pierre Perron (1991) "Pitfalls and Opportunities: What Macroeconomists Should Know about Unit Roots," *NBER Macroeconomics Annual*, 141–201.
- Chambers, John M., William S. Cleveland, Beat Kleiner, Paul A. Tukey (1983) *Graphical Methods for*

- Data Analysis*, Wadsworth & Brooks/Cole Publishing Company.
- Chesher, A., T. Lancaster, and M. Irish (1985) "On Detecting the Failure of Distributional Assumptions," *Annales de L'Insee*, 59/60, 7–44.
- Chesher, A. and M. Irish (1987) "Residual Analysis in the Grouped Data and Censored Normal Linear Model," *Journal of Econometrics*, 34, 33–62.
- Cleveland, William S. (1993) *Visualizing Data*, Hobart Press.
- Cleveland, William S. (1994) *The Elements of Graphing Data*, Hobart Press.
- Conover, W.J. (1980) *Practical Nonparametric Statistics*, 2nd edition, John Wiley & Sons.
- Conover, W.J., M.E. Johnson and M.M. Johnson (1981) "A Comparative Study of Tests for Homogeneity of Variance with Applications to the Outer Continental Shelf Bidding Data," *Technometrics*, 23, 351–361.
- Davidson, Russell and James G. MacKinnon (1989) "Testing for Consistency using Artificial Regressions," *Econometric Theory*, 5, 363–384.
- Davidson, Russell and James G. MacKinnon (1993) *Estimation and Inference in Econometrics*, Oxford University Press.
- Dezhbaksh, Hashem (1990) "The Inappropriate Use of Serial Correlation Tests in Dynamic Linear Models," *Review of Economics and Statistics*, 126–132.
- Dickey, D.A. and W.A. Fuller (1979) "Distribution of the Estimators for Autoregressive Time Series with a Unit Root," *Journal of the American Statistical Association*, 74, 427–431.
- Doan, Thomas, Robert B. Litterman, and Christopher A. Sims (1984) "Forecasting and Conditional Projection using Realistic Prior Distributions," *Econometric Reviews*, 3, 1–100.
- Dyer, D.D. and J.P. Keating (1980) "On the Determination of Critical Values for Bartlett's Test," *Journal of the American Statistical Association*, 75, 313–319.
- Engle, Robert F. (1982) "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of U.K. Inflation," *Econometrica*, 50, 987–1008.
- Engle, Robert F. (1984) "Wald, Likelihood Ratio, and Lagrange Multiplier Tests in Econometrics," in Chapter 13 of *Handbook of Econometrics*, Volume 2, North-Holland.
- Engle, Robert F. and C.W.J. Granger (1987) "Co-integration and Error Correction: Representation, Estimation, and Testing," *Econometrica* 55, 251–276.
- Engle, Robert F., David M. Lilien, and Russell P. Robins (1987) "Estimating Time Varying Risk Premia in the Term Structure: The ARCH-M Model," *Econometrica* 55, 391–407.
- Engle, Robert F. and Victor K. Ng (1993) "Measuring and Testing the Impact of News on Volatility," *Journal of Finance*, 48, 1022–1082.
- Engle, Robert F. and Mark W. Watson (1987) "The Kalman filter: Applications to Forecasting and Rational-Expectations Models" in Chapter 7 of *Advances in Econometrics—Fifth World Congress*, Volume 1, Cambridge University Press.
- Evans, M., N. Hastings, and B. Peacock (1993) *Statistical Distributions*, 2nd edition, John Wiley & Sons.
- Fahrmeir, Ludwig, and Gerhard Tutz (1994) *Multivariate Statistical Modelling Based on Generalized Linear Models*, Springer.
- Fair, Ray C. (1970) "The Estimation of Simultaneous Equation Models With Lagged Endogenous Variables and First Order Serially Correlated Errors," *Econometrica*, 38, 507–516.
- Fair, Ray C. (1978) "A Theory of Extramarital Affairs," *Journal of Political Economy*, 86, 45–61.
- Fair, Ray C. (1984) *Specification, Estimation, and Analysis of Macroeconomic Models*, Harvard University Press.
- Fama, Eugene F. and Michael R. Gibbons (1982) "Inflation, Real Returns, and Capital Investment" *Journal of*

- Monetary Economics*, 9, 297–323.
- Fan, J. and Gijbels, I. (1996) *Local Polynomial Modelling and Its Applications*, Chapman & Hall.
- Fan, J. and J.S. Marron (1994) “Fast Implementations of Nonparametric Curve Estimators,” *Journal of Computational and Graphical Statistics*, 3, 35–56.
- Gerfin, M., “Parametric and semiparametric estimation of the binary response model of labour market participation,” *Journal of Applied Econometrics*, Vol. 11, No. 3, 1996, 321–340.
- Glosten, L.R., R. Jagannathan, and D. Runkle (1993) “On the Relation between the Expected Value and the Volatility of the Normal Excess Return on Stocks,” *Journal of Finance*, 48, 1779–1801.
- Godfrey, L.G. (1988) *Specification Tests in Econometrics*, Cambridge University Press.
- Goldberger, Arthur S. (1991) *A Course in Econometrics*, Harvard University Press.
- Gourieroux, C., A. Monfort, and C. Trognon (1984a) “Pseudo-Maximum Likelihood Methods: Theory,” *Econometrica*, 52, 681–700.
- Gourieroux, C., A. Monfort, and C. Trognon (1984b) “Pseudo-Maximum Likelihood Methods: Applications to Poisson Models,” *Econometrica*, 52, 701–720.
- Gourieroux, C., A. Monfort, E. Renault, and A. Trognon (1987) “Generalized Residuals,” *Journal of Econometrics*, 34, 5–32.
- Granger, C. W. J. (1969) “Investigating Causal Relations by Econometric Models and Cross-Spectral Methods,” *Econometrica*, 37, 424–438.
- Grasa, Antonio Aznar (1989) *Econometric Model Selection: A New Approach*, Kluwer.
- Greene, William H. (1997) *Econometric Analysis*, 3rd Edition, Prentice Hall.
- Gujarati, Damodar N. (1995) *Basic Econometrics*, 3rd Edition, McGraw-Hill.
- Hamilton, James D. (1994a) *Time Series Analysis*, Princeton University Press.
- Hamilton, James D. (1994b) “State Space Models” in Chapter 50 of *Handbook of Econometrics*, Volume 4, North-Holland.
- Härdle, Wolfgang (1991) *Smoothing Techniques with Implementation in S*, Springer-Verlag.
- Harvey, Andrew C. (1987) “Applications of the Kalman Filter in Econometrics” in Chapter 8 of *Advances in Econometrics—Fifth World Congress*, Volume 1, Cambridge University Press.
- Harvey, Andrew C. (1989) *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press.
- Harvey, Andrew C. (1990) *The Econometric Analysis of Time Series*, 2nd edition, MIT Press.
- Harvey, Andrew C. (1993) *Time Series Models*, 2nd edition, MIT Press.
- Hausman, Jerry A. (1978) “Specification Tests in Econometrics,” *Econometrica*, 46, 1251–1272.
- Hodrick, R.J. and E.C. Prescott (1997) “Postwar U.S. Business Cycles: An Empirical Investigation,” *Journal of Money, Credit, and Banking*, 29, 1–16.
- Hosmer, David W. Jr. and Stanley Lemeshow (1989) *Applied Logistic Regression*, John Wiley & Sons.
- Hylleberg, S. and G.E. Mizon (1989) “A Note on the Distribution of the Least Squares Estimator of a Random Walk with Drift,” *Economics Letters*, 29, 225–230.
- Johansen, Soren and Katarina Juselius (1990) “Maximum Likelihood Estimation and Inferences on Cointegration—with applications to the demand for money,” *Oxford Bulletin of Economics and Statistics*, 52, 169–210.
- Johansen, Soren (1991) “Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models,” *Econometrica*, 59, 1551–1580.

- Johansen, Soren (1995) *Likelihood-based Inference in Cointegrated Vector Autoregressive Models*, Oxford University Press.
- Johnson, Norman L. and Samuel Kotz (1969) *Discrete Distributions*, Houghton Mifflin.
- Johnson, Norman L. and Samuel Kotz (1970) *Continuous Univariate Distributions 1 & 2*, Houghton Mifflin.
- Johnston, Jack and John Enrico DiNardo (1997) *Econometric Methods*, 4th Edition, McGraw-Hill.
- Judge, George G., W.E. Griffiths, R. Carter Hill, Helmut Lutkepohl, and Tsoung-Chao Lee (1985) *The Theory and Practice of Econometrics*, 2nd edition, John Wiley & Sons.
- Kennan, John (1985) "The Duration of Contract Strikes in U.S. Manufacturing," *Journal of Econometrics*, 28, 5–28.
- Levene, H. (1960) "Robust Tests for the Equality of Variances," in I. Olkin, S.G. Ghurye, W. Hoeffding, W.G. Madow, and H.B. Mann (eds) *Contribution to Probability and Statistics*, Stanford University Press.
- Ljung, G. and G. Box (1979) "On a Measure of Lack of Fit in Time Series Models," *Biometrika*, 66, 265–270.
- Lutkepohl, Helmut (1991) *Introduction to Multiple Time Series Analysis*, Springer-Verlag.
- MacKinnon, J.G. (1991) "Critical Values for Cointegration Tests," Chapter 13 in *Long-run Economic Relationships: Readings in Cointegration*, edited by R.F. Engle and C.W.J. Granger, Oxford University Press.
- Marron, J.S. and D. Nolan (1989) "Canonical Kernels for Density Estimation," *Statistics and Probability Letters*, 7, 191–195.
- McCallum, Bennett T. (1989) "Real Business Cycle Models," Chapter 1 in *Modern Business Cycle Theory*, edited by Robert J. Barro, Harvard University Press.
- McCullagh, P. and J.A. Nelder (1989) *Generalized Linear Models*, 2nd Edition, Chapman & Hall.
- McDonald, J. and R. Moffitt (1980) "The Uses of Tobit Analysis," *Review of Economic and Statistics*, 62, 318–321.
- Nelson, Daniel B. (1991) "Conditional Heteroskedasticity in Asset Returns: A New Approach," *Econometrica*, 59, 347–370.
- Newey, Whitney and Kenneth West (1987a) "Hypothesis Testing with Efficient Method of Moments Estimation," *International Economic Review*, 28, 777–787.
- Newey, Whitney and Kenneth West (1987b) "A Simple Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708.
- Newey, Whitney and Kenneth West (1994) "Automatic Lag Selection in Covariance Matrix Estimation," *Review of Economic Studies*, 61, 631–653.
- Osterwald-Lenum, Michael (1992) "A Note with Quantiles of the Asymptotic Distribution of the Maximum Likelihood Cointegration Rank Test Statistics," *Oxford Bulletin of Economics and Statistics*, 54, 461–472.
- Pagan, A. and F. Vella (1989) "Diagnostic Tests for Models Based on Individual Data: A Survey," *Journal of Applied Econometrics*, 4, S29–S59.
- Phillips, Peter C.B. and S. Ouliaris (1990) "Asymptotic properties of residual based tests for cointegration," *Econometrica*, 58, 165–193.
- Phillips, P.C.B. and P. Perron (1988) "Testing for a Unit Root in Time Series Regression," *Biometrika*, 75, 335–346.
- Pindyck, Robert S. and Daniel L. Rubinfeld (1991) *Econometric Models and Economic Forecasts*, 3rd edition, McGraw-Hill.
- Powell, J.L. (1986) "Symmetrically Trimmed Least Squares Estimation for Tobit Models," *Econometrica*, 54, 1435–1460.
- Prescott, Edward C. (1986) "Theory Ahead of Business-Cycle Measurement," *Carnegie-Rochester Conference*

- Series on Public Policy*, 25, 11–44.
- Press, W.H., S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery (1992) *Numerical Recipes in C*, 2nd edition, Cambridge University Press.
- Quandt, Richard E. (1983) “Computational Problems and Methods,” Chapter 12 in *Handbook of Econometrics*, Volume 1, North-Holland.
- Ramsey, J. B. (1969) “Tests for Specification Errors in Classical Linear Least Squares Regression Analysis,” *Journal of the Royal Statistical Society, Series B*, 31, 350–371.
- Ramsey, J. B. and A. Alexander (1984) “The Econometric Approach to Business-Cycle Analysis Reconsidered,” *Journal of Macroeconomics*, 6, 347–356.
- Rao, P. and Z. Griliches (1969) “Small Sample Properties of Several Two-Stage Regression Methods in the Context of Auto-Correlated Errors,” *Journal of the American Statistical Association*, 64, 253–272.
- Said, Said E. and David A. Dickey (1984) “Testing for Unit Roots in Autoregressive Moving Average Models of Unknown Order,” *Biometrika*, 71, 599–607.
- Sheskin, David J. (1997) *Parametric and Nonparametric Statistical Procedures*, CRC Press.
- Silverman, B.W. (1986) *Density Estimation for Statistics and Data Analysis*, Chapman & Hall.
- Simonoff, Jeffrey S. (1996) *Smoothing Methods in Statistics*, Springer-Verlag.
- Tauchén, George (1986) “Statistical Properties of Generalized Method-of-Moments Estimators of Structural Parameters Obtained From Financial Market Data,” *Journal of Business & Economic Statistics*, 4, 397–416.
- Temme, Nico M. (1996) *Special Functions: An Introduction to the Classical Functions of Mathematical Physics*, John Wiley & Sons.
- Watson, Mark W. and Robert F. Engle (1983) “Alternative Algorithms for the Estimation of Dynamic Factor, MIMIC and Varying Coefficient Regression Models,” *Journal of Econometrics*, 23, 385–400.
- White, Halbert (1980) “A Heteroskedasticity-Consistent Covariance Matrix and a Direct Test for Heteroskedasticity,” *Econometrica*, 48, 817–838.
- White, Halbert (1982) “Maximum Likelihood Estimation of Misspecified Models,” *Econometrica*, 50, 1–26.
- Wooldridge, Jeffrey M. (1996) “Quasi-Likelihood Methods for Count Data,” forthcoming in the *Handbook of Applied Econometrics*.
- Zakoian, J.M. (1990) “Threshold Heteroskedastic Models,” manuscript, CREST, INSEE, Paris.





# Index

---

## **Symbols**

- + 102615 5
- + 120318 202
- .DB file 105
- .DB? file 105
  - automatic backup 580
- .EDB file 102
- .RTF file 226
- .WF1 file 36
- ? (pool cross section identifier) 527
- @all
  - in assign statement 555
  - with smpl 59, 585
- @exclude 557
- @first 59, 585
- @-functions
  - date functions 565
  - descriptive statistics 568
  - estimation results 241
  - integrals and other special functions 569
  - mathematical functions 566
  - p-values 575
  - statistical distribution functions 572
  - time series functions 567
  - trigonometric functions 572
- @last 59, 585
- @prefix 556
- @trace 557
- ~, in backup file name 580

## **Numerics**

- 2SLS 253
- 3SLS 457, 471

## **A**

- Active window 7, 45
- Add

- shade to graph 220
  - text to graph 219
- Add factors 554
- Adjusted R-squared 239, 242, 576
- Adjustment parameter 489
- Advanced query 122
- AIC 241
- Akaike criterion 241, 242, 386, 576, 599
- Alias
  - database 118
  - OBALIAS.INI file 135
  - object 134
- All (database) 103
- Almon lag 293
- Analysis of variance 151
  - by ranks 154
- and operator 60, 85, 565
- Andrews test 391, 429
- ANOVA 151
  - by ranks 154
- AR specification
  - estimation 277
  - forecast 320
  - in 2SLS 256
  - in ARIMA models 281
  - in models 553
  - in nonlinear 2SLS 266
  - in nonlinear least squares 264
  - in pool 538
  - in state space 510
  - in systems 459
- AR(1)
  - coefficient 274
  - Durbin-Watson statistic 274
  - estimation 278
- AR(p) 274
  - estimation 278
- Arc cosine 572

- Arc sine 572
- Arc tangent 572
- ARCH
  - ARCH-M model 361
  - asymmetric models 368
  - component models 373
  - correlogram test 337
  - LM test 339
  - mean equation 362
  - QML standard errors 363
  - standardized residual 368
  - starting values 380
  - Threshold ARCH (TARCH) 368
- ARIMA models 281
  - Box-Jenkins approach 282
  - diagnostic checking 292
  - difference operator 284
  - identification 282
  - specification 283
  - starting values 289
  - state space representation 510, 512
- ARMA terms 285
  - in models 553
  - in solve 559
  - seasonal 286
  - testing 338
- Artificial regression 340, 351
- ASCII
  - file export 69
  - file import 68, 78
- Assign 555
- Asymmetric (G)ARCH models 368
  - component model 374
- Asymptotic test 329
- Augmented Dickey-Fuller test 298
- Augmented regression 344
- Auto choice 217
- Autocorrelation 240
- Autocorrelation function 157
- Autoregressive conditional heteroskedasticity 359
- Auto-search
  - database 119
- Auto-series
  - generate new series 93, 116
  - in estimation 96
  - in groups 95
  - in regression 116
  - with database 116
- Autospec 507
- Auxiliary regression 338, 339
- Average log likelihood 386
- Axes display options 220
- B**
- Backcast 290
  - in models 553
- Backup file option 580
- Balanced data 532
- Balanced sample 537
- Bandwidth 202
  - Andrews 474
  - bracket option 202, 208
  - in kernel fit 211
  - in nearest neighbor fit 208
  - Newey-West (fixed) 474
  - Newey-West (variable) 475
  - selection in GMM 461, 474
- Bar graph 216
  - options 218
- Bartlett kernel 473
- Bartlett test 155
- Behavioral equation 552
- Berndt-Hall-Hall-Hausman 593
- Beta distribution 573
- Beta integral 570
- BHHH 593
- Bias proportion 317
- Binary dependent variable 381
  - error messages 388
  - fitted index 394
  - interpretation of coefficient 385
  - log likelihood 382
- Binning option 146, 189
  - exact 203, 211

linear 203, 211  
 Binomial coefficient 570  
 Binomial distribution 573  
 Binomial sign test 150  
 Box around graph 217  
 Box-Cox transformation 206  
 Box-Jenkins 282  
 Bracket bandwidth option 202  
   in kernel fit 212  
   in nearest neighbor fit 208  
 Breakpoint Chow test 341  
 Breusch-Godfrey test 275, 339  
 Brown-Forsythe test 155  
 Button bar option 580  
**C**  
 C (coef vector) 232  
 C (constant) 232  
 Cache 132  
 Capitalization 48  
 Categorical regressor stats 389, 410  
 Causality  
   Granger's test 194, 486  
 cd (change directory) 38  
 Censored dependent variable 405  
   fitted index 411  
   goodness-of-fit tests 412  
   interpretation of coefficient 410  
   log likelihood 406  
   scale factor 410  
 Censoring point 407  
   unknown 408  
 Census X-11 162  
   limitations 162  
 Centered seasonal dummy 491  
 Chi-square distribution 573  
 Chi-square statistic 332  
 Chi-square EMBED Word.Picture.6 - test  
   for the median 154  
   n-way table 191  
 Cholesky decomposition 481

Chow test  
   breakpoint 341  
   forecast 342  
   n-step forecast 349  
   one-step forecast 348  
 Classification table  
   binary models 389  
   ordered models 403  
   sensitivity 390  
   specificity 390  
 Cleveland subsampling 208  
 Click(ing) 6  
 Clipboard 224, 226  
 Close  
   EViews 13  
   object 580  
   window 7  
 Cochrane-Orcutt 257, 280  
 Coef (coefficient vector)  
   default 232  
   update 244  
 Coefficient  
   common (pool) 538  
   covariance matrix 242  
   cross-section specific (pool) 538  
   estimated 237  
   from a VAR 486  
   from a VEC 496  
   recursive estimates 349  
   restrictions 234  
   saved vector 242  
   standard error 238  
 Coefficient restriction test 330  
 Cointegrating rank 489  
 Cointegrating vector 490  
   estimated 494  
   normalization 490, 494  
   standard errors 503  
 Cointegration 487  
   number of relations 489  
   trend assumption 490  
 Cointegration test 489

- carrying out the test 493
- critical values 490
- maximum eigenvalue statistic 494
- summary test 495
- trace statistic 493
- trend assumption 490
- Collinearity 248, 293
- Column width 225
- Command window 12
- Comments 40, 48, 555
- Common coefficients (pool) 538
- Common sample 87
- Complementary log-log function 570
- Component (G)ARCH 373
  - asymmetric component 374
  - permanent component 373
  - transitory component 373
- Conditional standard deviation (SD)
  - graph 365
- Conditional variance 357, 359
  - forecast 366
  - in the mean equation 361
  - make series 366
- Constant 232, 237
  - in ordered models 401
  - in pool estimation 538
- Contingency coefficient 191
- Contract range 38
- Convergence criterion 596
  - in model solve 561
  - in nonlinear least squares 263, 267
  - in pool estimation 540
- Coordinates 219
- Copy
  - and paste 49
  - data 68
  - database 127
  - datadata\_cutpaste 64
  - graph 224
  - object 49
  - table 225
  - to and from database 108
- Correlation 569
- Correlogram 159, 275
  - autocorrelation function 157
  - cross 192
  - partial autocorrelation function 158
  - Q-statistic 159
  - squared residuals 337, 365
- Cosine 572
- Count models 418
  - negative binomial (ML) 420
  - Poisson 419
- Covariance 569
- Covariance matrix
  - element 242, 576, 577
  - HAC (Newey-West) 252
  - heteroskedasticity consistent 251
  - of estimated coefficients 242
- Covariance proportion 317
- Cramer's V 191
- Create
  - database 103
  - group 57, 96
  - object 43
  - series 53, 88
  - text object 226
  - workfile 34
- Cross correlation 192
  - test asymmetry 372
- Cross correlogram 192
  - test asymmetry 372
- Cross section identifiers 526
- Cross-equation
  - coefficient restriction 456, 458
  - correlation 457
  - weighting 456
- Cross-section specific coefficients (pool) 538
- Cumulative distribution 197
  - Blom option 198
  - Ordinary option 198
  - Rankit option 198
  - Tukey option 198
  - Van der Waerden option 198

CUSUM of squares test 347

CUSUM test 346

## D

Dashed line in graph 217

Data

cut and paste 68

cut and pastedata\_cutpaste 66

exportdata\_export 70

keyboard entrydata\_keyboard 63

missing 88

pool 527

Database

alias 118

auto-search 119

auto-series 117

cache 132

copy 127

copy objects 108

create 103

default 104

delete 127

delete objects 111

DRI Basics 129

export 107

fetch objects 106

field 122

Haver Analytics 130

link 101, 131

open 103

packing 127

queries 120

rebuild 129

remote DRI databases 130

rename 127

rename object 110

repair 128

sharing violation 104

statistics 128

store objects 105

test integrity 128

window 103

Database registry 118

Date

format 583

functions 565

notation option 581

Dated data table 174

formatting options 179

frequency conversion 178

row options 179

table options 175

transformation methods 177

Default

database 12, 104

directory 12

equation 36, 245

graph options 218

set directory 37

update directory 37

Default database in search order 119

Delete 50

cells from table 225

database 127

from pool 536

objects from database 111

Delete observation 56

Demonstration 30

creating a workfile 15

estimation 22

forecasting 27

import data 18

specification test 24

Description field 126

Descriptive Statistics

series 147

Descriptive statistics

@-functions 568

balanced sample (pool) 534

by classification 145

common sample (group) 187

common sample (pool) 534

cross-section specific 534

group 187

individual samples (group) 187

individual samples (pool) 533

- stacked data 534
- time period specific 534
- Deselect all 44
- Dialog box 9
- Difference from moving-average 163
- Difference operator 86, 284
  - seasonal 86, 284
- Display filter 36, 40
- Display\_name field 126
- Distribution graphs 155, 197
  - cumulative distribution 197
  - kernel density 156, 201
  - QQ-plot 156, 199
  - quantile 198
  - survivor 197
- Drag(ging) 6
  - legend 221
  - text in graph 220
- DRI
  - Basic Economics Database 129
  - frequency 136
  - illegal names 134
  - object alias 134
  - queries 136
  - remote access 130
  - shadowing 135
  - troubleshooting 137
- Dummy variable 92
  - censoring point 408
  - dependent 381
  - pool series 535
  - seasonal 568
- Durbin-Watson statistic 240, 242, 274, 576
  - lagged dependent variable 275
- Dynamic
  - forecast 307
  - solution 558

## **E**

- Easy query 120
- Edit
  - group 58
  - series 54
  - table 225
- End field 125
- Endogeneity 351
- Endogenous variable 253
- Epanechnikov kernel 202, 210
- Equality tests 151
  - mean 151
  - median 153
  - variance 154
- Equation
  - constant 232
  - create 231
  - default 245
  - retrieve 246
  - specification 232
  - store 246
- Error correction 487
- Error function 571
- Error message display option 580
- Estimation
  - AR specification 277
  - auto-series 96
  - binary dependent variable 383
  - censored models 406
  - collinearity 248
  - count models 418
  - for pool 536
  - missing data 235
  - nonlinear least squares 260
  - options 594
  - ordered models 400
  - output 237
  - residuals 245
  - sample 235
  - sample adjustment 236
  - saved results 241
  - single equation methods 235
  - state space 514
  - systems 460
  - truncated models 415
  - two stage least squares 253

- weighted least squares 251
- Euler's constant 571
- EViews Databases 101
- Exogenous variable 253
- Expand range 38
- Expectation-prediction table
  - binary models 389
  - ordered models 403
- Expected dependent variable
  - censored models 411
  - truncated models 417
- Expected latent variable
  - censored models 411
  - truncated models 417
- Exponential distribution 574
- Exponential function 567
- Exponential GARCH (EGARCH) 370
- Exponential smoothing 169
  - double 165
  - Holt-Winters additive 166
  - Holt-Winters multiplicative 165
  - Holt-Winters no-seasonal 167
  - single 164
- Export data
  - from pool objects 533
  - to ASCII files 69
  - to spreadsheet files 69
- Export database 107
- Export datadata\_export 70
- Expression 81
  - for database fields 123
  - logical 84
  - parentheses 82
- Extended search 560
- Extreme value
  - binary model 384
- Extreme value distribution 574

## F

- Factorial 567
- Failure to improve 597
- F-distribution 574
- Fetch
  - from database 106
  - from pool 536
- Fields in database 122
  - description 126
  - display\_name 126
  - end 125
  - expressions 123
  - freq 124
  - history 126
  - last\_update 125
  - last\_write 125
  - name 123
  - remarks 126
  - source 126
  - start 125
  - type 124
  - units 126
- Finite sample distribution 329
- First order serial correlation 274
- Fitted index
  - binary models 394
  - censored models 411
  - truncated models 417
- Fitted probability
  - binary models 393
- Fixed
  - decimal 225
  - digits 225
- Fixed effects 539, 546
  - standard errors 547
- Fixed variance parameter
  - negative binomial QML 422
  - normal QML 422
- followed 552
- Font options 579
  - axes 220
  - graph 218
  - legend 221
  - table 225
  - text in graph 219
- Force scale through zero 217

## Forecast

- AR specification 320
- binary models 393
- by exponential smoothing 169
- censored models 411
- Chow test 342
- conditional variance 366
- count models 423
- dynamic 307
- equations with formula 323
- error 313
- evaluation 315
- fitted values 312
- from estimated equation 305
- interval 315
- MA specification 321
- n-step test 349
- one-step test 348
- ordered models 404
- out-of-sample 312
- standard error 314, 325
- static 307
- structural 307, 320
- truncated models 417
- variance 313
- with models 562

## Formula

- forecast 323
- implicit assignment 90
- normalize 91
- specify equation 233

## Freeze 50

- output 582

## Freq field 124

## Frequency 583

## Frequency conversion 49, 111

- DRI database 136
- in dated data table 178
- propagate NAs 112
- undated series 112

## F-statistic 241, 242, 332, 335, 576

## F-test

- variance equality 154

## Full information maximum likelihood 457

**G**

## Gamma distribution 574

## Gamma integral 571

## GARCH

- asymmetric models 368
- component models 373
- exponential GARCH (EGARCH) 370
- GARCH(1,1) model 359
- GARCH(p,q) model 361
- mean equation 362
- news impact curve 371
- QML standard errors 363
- standardized residual 368
- starting values 380
- Threshold ARCH (TARCH) 368

## Gaussian distribution 575

## Gauss-Newton 593

## Gauss-Seidel 559, 597

## Generalized ARCH 359

## Generalized error distribution 370

## Generalized least squares 539

## Generalized linear models 428

- quasi-likelihood ratio test 422
- robust standard errors 428
- variance factor 428

## Generalized method of moments 457

## Generalized residual

- binary models 394
- censored models 411
- count models 423
- ordered models 405
- score vector 395
- truncated models 416

## Generate series 88

- by command 91
- dynamic assignment 90
- for pool 535
- implicit formula 90
- using samples 89

## Genr 535



- 
- Geometric moving average 95
  - GLM 428
  - GLM standard errors 428
  - GLS 539
  - GMM
    - bandwidth selection 461
    - for systems 457
    - HAC weighting matrix 473
    - J-statistic 270
    - kernel options 461
    - orthogonality condition 268
    - overidentifying restrictions 270
    - prewhitening option 461, 475
    - single equation 267
    - system 472
    - White weighting matrix 473
  - Gompit 384
  - Goodness-of-fit
    - adjusted R-squared 239
    - Andrews test 391, 429
    - forecast 315
    - Hosmer-Lemeshow test 391, 428
    - R-squared 239
  - Granger causality 194, 486
  - Graph attributes 217
  - Graph objects
    - add shade 220
    - add text 219
    - axes 220
    - legends 221
    - position text 219
  - Graph print 223
  - Graph scale
    - force through zero 217
    - log scale 217
    - manual 218
    - normalize 217
    - optimized 217
  - Graph template 221
  - Graph type 216
    - bar 216
    - high-low 217
    - line 216
    - mixed bar and line 216
    - pie chart 217
    - scatter diagram 216
    - special 216
    - stacked bars 216
    - stacked lines 216
    - XY-line 217
  - Graph views
    - default 218
    - options 215
  - Graphs
    - merge 44
    - options 222
  - Grid
    - in table 225
    - lines in graph 217
  - Grid search 594
  - Group 98
    - add member 173
    - auto-series 95
    - create 57, 96
    - editing 58
    - element 97
    - spreadsheet view 173, 581
  - Group into bins option 146, 189
  - Groupsec\_group 58
  - Groupwise heteroskedasticity 187
- ## H
- Hannan-Quinn criterion 386, 599
  - Hatanaka two step estimator 280
  - Hatching 217
  - Hausman test 351
  - Haver Analytics Databases 130
  - Help
    - help system 13
    - World Wide Web 14
  - Heteroskedasticity
    - binary models 398
    - groupwise 187
    - of known form 249

- of unknown form 251
- White's test 340
- Heteroskedasticity and autocorrelation consistent covariance 252
- Heteroskedasticity consistent covariance matrix 251
- High-low-close 183, 217
- Hildreth-Lu 280
- Histogram 144, 338
- History field 126
- Hodrick-Prescott filter 170
- Holt-Winters
  - additive 166
  - multiplicative 165
  - no-seasonal 167
- Horizontal line in table 225
- Hosmer-Lemeshow test 391, 428
- Huber/White standard errors 427
- Hypothesis tests 329
  - ARCH 337, 339
  - Bartlett test 155
  - binomial sign test 150
  - Brown-Forsythe 155
  - chi-square test 154
  - CUSUM 346
  - CUSUM of squares 347
  - F-test 154
  - Hausman test 351
  - heteroskedasticity 340
  - irrelevant variable 336
  - Kruskal-Wallis test 154
  - Levene test 155
  - Mann-Whitney U-test 153
  - multi-sample 151
  - nonnested 353
  - normality 338
  - omitted variables 335
  - p-value 238
  - Siegel-Tukey test 155
  - single sample 148
  - stability 341
  - Van der Waerden test 150, 154
  - Wald coefficient restriction test 330
  - Wilcoxon signed ranks test 150
- I**
- Icon 42
- Identification
  - Box-Jenkins 282
  - nonlinear models 267
  - state space models 523
- Identity
  - in models 554
  - in system 459
- If condition in samples 60
- Import data
  - for pool objects 527, 532
  - from ASCII 68
  - from ASCIIdata\_asciiread 70
  - from spreadsheet 66
- Impulse response 481, 502
  - Monte Carlo standard errors 503
- Incomplete beta integral 570
- Incomplete gamma integral 571
- Incorrect functional form 340, 344
- Index
  - fitted from binary models 394
  - fitted from censored models 411
  - fitted from truncated models 417
- Individual sample 87
- Information criterion
  - Akaike 241, 599
  - Hannan-Quinn 599
  - Kullback-Leibler 599
  - Schwarz 241, 599
- Initial values 516
- Inner product 569
- Innovation 273
- Insert
  - cells to table 225
  - observation 56
- Insertion point 11
- Instrumental variable 253
  - for 2SLS with AR specification 257
  - for GMM 268

- for nonlinear 2SLS 265
- identification 460
- in systems 459
- order condition 254

Instruments 253

Integer dependent variable 418

Integrated series 297

Integrity (database) 128

Intercept 232, 237

- in pool estimation 538

Inverted AR roots 280, 287

Inverted MA roots 287

Irrelevant variable test 336

Iteration option 596

- for pool estimation 539

- for system estimation 462

- in model solve 560

- in nonlinear least squares 263

- state space 517

## J

Jarque-Bera statistic 145, 338, 366

Johansen's cointegration test 489

J-statistic 270

- saved value 242, 576

J-test 353

Justification 219, 225

## K

Kalman filter 506

- iteration algorithm 522

- starting values 522

Kernel density 156, 201, 210

Kernel fit 209

Kernel function

- in density estimation 207

- in kernel regression 210

Kernel option 461, 473

- Bartlett 473

- quadratic spectral 474

Keyboard

- data entrydata\_keyboard 63

- focus option 579

- shortcuts 9

Kruskal-Wallis test 154

Kullback-Leibler 599

Kurtosis 145

## L

Label 40, 49

- automatic update option 580

- capitalization 48

- pie graph 184, 218

- series 160

Lag

- dynamic assignment 90

- series 85

Lagged dependent variable

- and serial correlation 273

- Durbin-Watson statistic 275

Lagrange multiplier test 275

Laplace distribution 574

Large sample test 329

Last\_update field 125

Last\_write field 125

Latent variable

- binary model 382

- censored models 406

- ordered models 399

Lead

- series 85

Legend options 221

Levene test 155

Leverage effect 369

Likelihood 240

Likelihood ratio test 334

Limit points 402

- make covariance matrix 404

- make vector 404

- non-ascending 403

Limited dependent variable 381

Line graph 216

- options 218

Line patterns 218  
Link 101, 131  
    options 133  
Ljung-Box Q-statistic 159  
    serial correlation test 275  
LM test  
    ARCH 339  
    auxiliary regression 338  
    serial correlation 275, 338  
Load  
    workfile 37  
Local regression 207  
Local weighting option 208  
LOESS 207  
loess 209  
Log likelihood  
    average 386  
    censored models 406  
    exponential 421  
    for binary models 382  
    negative binomial 420  
    normal 240, 422  
    ordered models 400  
    Poisson model 419  
    restricted 386  
    saved value 242, 576, 577  
    truncated models 415  
Logarithmic scale 217  
Logical expression 84  
    in easy query 121  
Logistic 572  
Logistic distribution 574  
Logit 384  
Log-normal distribution 574  
Long name 48  
    for series 160  
LOWESS 207  
LR statistic 334, 335, 386

## **M**

MA specification  
    backcasting 290

    forecast 321  
    in ARIMA models 281  
    in models 553  
    in state space 510  
    in two stage least squares 259  
Mann-Whitney U-test 153  
Manual scaling 218  
Marginal significance level 238, 329  
Marquardt 593  
match operator 124  
Mathematical functions 566  
Maximize window 7  
Maximum 569  
Maximum eigenvalue statistic 494  
Maximum likelihood  
    full information 457  
    quasi-generalized pseudo-maximum likelihood 425  
    quasi-maximum likelihood 421  
McFadden R-squared 386  
Mean 144, 569  
Mean absolute error 316  
Mean absolute percentage error 316  
Measurement equation 506  
Measurement error 253, 344  
Median 144, 569  
Menu 9, 46  
    main 9, 11  
    objects 47  
Merge  
    equations into a model 551  
    graphs 44  
    store option 106  
Metafile 224  
Micro TSP 5  
    workfiles 38  
Minimize window 7  
Minimum 569  
Missing data 86  
    estimation 235  
    in frequency conversion 112  
Missing values

mathematical functions 566  
Mixed bar and line graph 216

## Model

- add factors 554
- ARMA specification 553
- assign statement 555
- comment 555
- create 551
- definition 456
- exclude statement 557
- extended search 560
- forecast 562
- identity 554
- name solved series 555
- nonlinear specification 553
- ordering of equations 559
- solve 558
- specify 551
- trace statement 557

## Mouse 6

Move window 8

Moving average 568  
geometric 95

Moving sum 568

Multicollinearity 248

## N

NA 86

- recode 567

Nadaraya-Watson 210

## Name

- capitalization 48
- display option 40
- equation 245
- object 47
- reserved 48

Name field 123

Near singular matrix

- binary models 388
- collinearity 247
- nonlinear models 267, 596
- polynomial distributed lag 293

pool estimation 539

RESET test 345

Nearest neighbor fit 207

Negative binomial distribution 575

Newey-West

- HAC covariance 252

- truncation lag 252

News impact curve 371

No cross terms 340

Noninvertible MA process 287, 291, 554

Nonlinear coefficient restriction

- LR test 334

- Wald test 334

Nonlinear least squares 260

- convergence criterion 263

- forecast standard errors 315

- iteration option 263

- specification 261

- starting values 262

- two stage 265

- two stage with AR specification 266

- weighted 264

- weighted two stage 266

- with AR specification 264, 278

Nonnested tests 353

Nonstationary time series 297

Normal distribution 575

- test for 338

Normalization 490, 494

- in models 553

Normalize formula 91

Normalized scale 217

N-step forecast test 349

Null hypothesis 329

Number format 225

N-way table 192

- chi-square tests 190

## O

Object 33

- basics 41

- copy 49

- create 43
- data 41
- delete 50
- freeze 50
- icon 42
- label 49
- naming 48
- open 44
- print 50
- procedure 42
- samplesec\_sampleobject 62
- show 45
- store 50
- type 42
- view 42
- window 45
- Objects menu 47
- Observation equation 506
- OLS (Ordinary least squares)
  - t-statistic 238
- OLS (ordinary least squares)
  - adjusted R-squared 239
  - coefficient standard error 238
  - coefficients 237
  - regression standard error 239
  - R-squared 239
  - sum of squared residuals 239
  - system estimation 456, 469
- Omitted variable test 335, 344
- One-step forecast test 348
- One-way tabulation 156
- Open
  - database 103
  - multiple objects 44
  - object 44
  - options 38
  - workfile 37
- Operator 81, 565
  - arithmetic 81
  - comparison 84
  - conjunction (and, or) 85
  - difference 86
  - lag 85
  - lead 85
  - parentheses 82
- Optimization algorithms 592
  - BHHH 593
  - convergence criterion 596
  - Gauss-Newton 593
  - grid search 594
  - iteration control 596
  - Marquardt 593
  - quadratic-hill climbing 592
  - starting values 595
  - step size 594
  - used in EViews 598
- Optimized scale 217
- Option settings
  - allow only one untitled 580
  - backup files 580
  - button bar 580
  - date notation 581
  - error message display 580
  - fonts 579
  - keyboard focus 579
  - output redirection 582
  - print setup 582
  - program execution mode 581
  - series auto label 580
  - spreadsheet view option 581
  - warn on close 580
- or operator 60, 85, 565
- Order condition for identification 254
- Order of stacked data 530
- Ordered dependent variable 399
  - error messages 403
  - log likelihood 400
- Ordinary least squares 237
- Ordinary residual
  - binary models 394
  - censored models 410
  - count models 423
  - truncated models 416
- Orthogonality condition 268, 472
- Orthogonalized

seasonal dummy 491  
 shocks 481, 502  
 Output redirection 582  
 Overdispersion 420, 428  
   specification test 423  
 Overidentifying restrictions 270

## P

Pack database 127  
 Packable space 104, 128  
 Panel data 525  
 Param (command) 262, 460, 596  
 Parameter constancy test 341  
 Pareto distribution 575  
 Parks estimator 539, 549  
 Partial autocorrelation 282  
   function 158  
 Paste 49  
   existing series 65  
   new series 64  
 PDL 292  
   far end restriction 294  
   forecast standard errors 315  
   near end restriction 293  
   specification 294  
 Pearson chi-square 190  
 Percentage change 568  
 Phi coefficient 191  
 Phillips-Perron test 300  
 pi (constant) 572  
 Pie graph 184, 217  
   options 218  
 Poisson distribution 575  
 Polynomial distributed lags 315  
 Pool  
   ? 527  
   AR specification 538  
   balanced data 532, 534  
   balanced sample 537  
   base name 526  
   coefficient test 544

common coefficients 538  
 convergence criterion 540  
 copy 526  
 create 526  
 cross section identifiers 526  
 cross-section specific coefficients 538  
 cross-section weighting 548  
 dummy variable 535  
 export data 533  
 fixed effects 539, 546  
 generate series 535  
 import data 527  
 iteration option 539  
 make group 536  
 make system 544  
 naming series 526  
 order 530  
 random effects 539, 547  
 residuals 544  
 series 527  
 stacked data 528  
 SUR weighting 549  
 unstacked data 528  
 White heteroskedasticity covariance 540, 549  
 workfile 525  
 PoolGenr 535  
 Position of text in graph 219  
 Postscript file 224  
 Power (raise to) 565  
 Prais-Winsten 280  
 Precedence of Evaluation 82  
 Predetermined variable 253  
 Prediction table  
   binary models 389  
   ordered models 403  
 Prewhitening 461, 475  
 Print 50  
   redirection 582  
   selected 50  
 Print graph 223  
   as postscript file 224  
   in color 223

- Print setup option 582
- Probability response curve 395
- Probit 383
- Procedure 42
- Program
  - execution option 581
- p-value 238, 329, 334, 575

## **Q**

- QML 421
- QML standard errors
  - GARCH models 363
- QQ-plot (quantile-quantile) 156, 199
  - multiple 186
- Q-statistic
  - Ljung-Box 159
  - serial correlation test 275
- Quadratic spectral kernel 474
- Quadratic-hill climbing 592
- Qualitative dependent variable 381
- Quantile 198
- Quasi-generalized pseudo-maximum likelihood 425
- Quasi-likelihood ratio test 422, 426
- Quasi-maximum likelihood 421
  - robust standard errors 427
- Queries on database 120
  - advanced query 122
  - DRI 136
  - easy query 120
  - examples 126
  - logical expressions 121
  - wildcard characters 120
- Quiet mode 581

## **R**

- Random coefficient model 507
- Random effects 539, 547
- Random walk 297
- Randomize ties 392
- Rank condition for identification 254
- Ratio to moving-average 162
- Read 527
- Rebuild database 129
- Recursive coefficient 349
  - save as series 349
- Recursive estimation
  - least squares 345
  - state space 508, 511
- Recursive residual 346
  - CUSUM 346
  - CUSUM of squares 347
  - n-step forecast test 349
  - one-step forecast test 348
  - save as series 350
- Redirect printer output 582
- Reduced rank 489
- Redundant variable test 336
- Registry 118
- Regression
  - adjusted R-squared 239
  - coefficient standard error 238
  - coefficients 237
  - collinearity 248
  - forecast 305
  - F-statistic 241, 576
  - log likelihood 240
  - residuals 245
  - results 237
  - R-squared 239
  - saved results 241
  - standard error 239
  - sum of squared residuals 239
  - t-statistic 238
- Remarks field 126
- Remove 220, 221
- Rename 47
  - database 127
  - objects in database 110
- Repair database 128
- Reparametrization 334
- Representation view
  - single equation 243
- Requirements 5



- 
- Reserved names 48
  - RESET test 344
  - Resid (default series) 245
  - Residual
    - covariance (pairwise) 577
    - from estimated equation 245
    - from two stage least squares 255
    - generalized 394, 411, 416, 423
    - make series 244
    - ordinary 394, 410, 416, 423
    - recursive 346
    - Standardized 394
    - standardized 244, 411, 416, 423
    - sum of squares 239
    - symmetrically trimmed 413
    - unconditional 274, 279
  - Residual-based tests 337
  - Resize window 8
  - Restore window 7
  - Restricted log likelihood 334, 386
  - Rich text format 226
  - Robust standard errors
    - GLM 428
    - Huber/White (QML) 427
  - Robustness iterations
    - with nearest neighbor fit 209
    - with regression 206
  - Root mean squared error 316
  - R-squared 239, 242
    - adjusted 239, 576
    - from two stage least squares 256
    - McFadden 386
    - saved 576
    - uncentered 339, 340
    - with AR specification 280
- S**
- Sample
    - @all 59, 585
    - @first 59, 585
    - @last 59, 585
    - adjustment in estimation 236
    - balanced 537
    - change 59
    - command 62
    - common 87
    - current 36
    - estimation 235
    - if condition 60
    - individual 87
    - objectsec\_sampleobject 62
    - with expressions 59, 585
    - workfile 58
  - Samplesec\_sample 61
  - SAR specification 286
  - Save
    - options 38
    - saveas 37
    - workfile 36
  - Scale factor 410
  - SCATMAT 184
  - Scatter diagrams 182, 205, 216
    - matrix 184
    - multiple 184
    - options 218
    - simple 182, 205
    - with kernel fit 209
    - with nearest neighbor fit 207
    - with regression 182, 205
  - Schwarz criterion 241, 242, 386, 576, 599
  - Score vector 395
  - Scroll window 7
  - Seasonal
    - ARMA terms 286
    - difference 86, 284
  - Seasonal adjustment 163
    - additive 163
    - census X-11 162
    - multiplicative 162
  - Seasonal dummy variable 568
    - centered 491
  - Seemingly unrelated regression 457
  - Select
    - all 44

- multiple items 8
- object 44
- single item 8
- sensitivity 390
- Serial correlation
  - ARIMA models 281
  - Durbin-Watson statistic 240, 274
  - nonlinear models 278
  - tests 274, 338
  - theory 273
  - two stage regression 279
- Series
  - bar graph view 143
  - create 53, 88
  - delete observation 56
  - descriptive statistics 144
  - difference 86
  - editing 54
  - extract observation 566
  - in pool objects 527
  - insert observation 56
  - label 160
  - lag 85
  - lead 85
  - line graph view 143
  - smpl + /- 55
  - spreadsheet view 143, 581
- Seriessec\_series 56
- Shade in graph 220
- Shadowing 135
- Sharing violation 104
- Show 44
- Siegel-Tukey test 155
- Simple hypothesis tests 148
  - mean 148
  - median 150
  - variance 149
- Simple scatter plot 182, 205
- Simultaneous equations 551
- Sine 572
- Skewness 145
- SMA specification 286
- Smoothing
  - methods 163
  - parameters 164
- Smpl command 62
- Smpl + /- 55
- Solve 558
  - algorithm 559
  - ARMA terms 559
  - dynamic 558
  - extended search 560
  - fit each equation 559
  - Gauss-Seidel 597
  - static 559
- Sort 38
- Source field 126
- Sparse label option 146, 189
- Specification test
  - for binary models 398
  - for overdispersion 423
  - for tobit 413
  - RESET (Ramsey) 344
  - White 340
- specificity 390
- Specify equation 232, 329
  - by formula 233
  - by list 232
  - in model 551
  - in systems 458
  - lagged series 233
  - nonlinear 261
  - with coef vector 235
  - with restrictions 234
  - without dependent variable 234
- Spreadsheet
  - file export 69
  - file import 66
  - view option 581
- Stability test 341
  - Chow breakpoint 341
  - Chow forecast 342
  - RESET 344
  - with unequal variance 350

- Stacked data 528
  - balanced 532
  - descriptive statistics 534
  - order 530
- Stacked graph
  - bar 216
  - line 216
- Standard deviation 144, 569
- Standard error
  - estimated coefficient 238
  - for fixed effects 547
  - forecast 314, 325
  - of regression 242, 576
  - saved vector 242
- standard error
  - of regression 239
- Standardized residual 244, 368
  - binary models 394
  - censored models 411
  - count models 423
  - state space 518
  - truncated models 416
- Start field 125
- Starting values 595
  - (G)ARCH models 380
  - binary models 387
  - for ARMA estimation 289
  - for nonlinear least squares 262
  - for systems 460
  - Param 262, 596
  - state space 516, 522
  - user supplied 289
- State equation 506
- State space 505
  - @mprior 517
  - @vprior 517
  - ARMA estimation 510
  - estimation 514
  - general specification 511
  - identification 523
  - iteration algorithm 522
  - Kalman filter 506
  - observation equation 506
  - recursive estimation 508
  - representation 505
  - smoothed state variable 518
  - standardized residual 518
  - starting values 522
  - state equation 506
  - time varying parameter 507
- State variable 505
  - one-step ahead 518
  - smoothed 518
- Statement in models 555
  - @exclude 557
  - @trace 557
  - assign 555
- Static
  - forecast 307
  - solution 559
- Stationary time series 297
- Statistical distribution functions 572
- Status line 12
- Step size 594
- Store 50
  - as .DB? file 106
  - from pool 536
  - in database 105
  - merge objects 106
- Structural option 307, 320, 559
- Sum of squared residuals 239, 242, 577
- Sum of squares 569
- Summation 569
- SUR 457, 470
  - in pool estimation 539, 549
- Survivor function 197
- Symmetrically trimmed residuals 413
- System
  - create 458
  - definition 455
  - estimation 456
  - estimation methods (technical) 468
  - specification 458

Table 224  
  as unformatted text 226  
  column width 225  
  copy to other windows programs 225  
  delete cells 225  
  edit 225  
  font 225  
  horizontal line 225  
  insert cells 225  
  number format 225  
  objects 224  
  title 225  
  views 224

Tabulation  
  n-way 188  
  one-way 156

Tangent 572

t-distribution 575

Template 221

Test statistic 329

Tests 329

Tests for structural change 341

Tests of equality 151

Text (object) 226

Theil inequality coefficient 316

Three stage least squares 457, 471

Threshold ARCH (TARCH) 368  
  component model 374  
  forecast 369

Ticks 220

Time trend 568

Time varying parameter 507  
  general specification 513

Title 225

Title bar 10, 36, 46

to (lag range) 233

Tobit 406

Toggle 9

Toolbar 36, 46

Trace statistic 493

Transition equation 506

Transpose 173

Trend series 92, 568

Trigonometric functions 572

Truncated dependent variable 415  
  fitted index 417  
  log likelihood 415

Truncation point 416

t-statistic 238  
  saved vector 242

Two stage least squares 259  
  nonlinear 265  
  nonlinear with AR specification 266  
  order condition 254  
  rank condition 254  
  residuals 255  
  system estimation 471  
  weighted 256  
  weighted in systems 457, 471  
  weighted nonlinear 266  
  with AR specification 256, 279  
  with MA specification 259

Type field 124

## **U**

Unconditional residual 279

Uniform distribution 575

Unit root test 298  
  augmented Dickey-Fuller 298  
  critical values 298, 302  
  how to carry out 302  
  Phillips-Perron 300  
  trend assumption 299

Units field 126

Unrestricted VAR 479

Unstacked data 528

Untitled 47, 48  
  allow only one option 580

Update  
  coefficient vector 244, 596  
  group 173

User supplied starting values 289

**V**

Van der Waerden test 150, 154

VAR

- decomposition 483
- estimation output 479
- impulse response 481
- lag order selection 498
- lag specification 478
- mathematical model 477
- orthogonalized shocks 481
- retrieving coefficients 486

Variance 569

Variance decomposition 483, 503

Variance factor 428

Variance proportion 317

VEC

- adjustment parameters 489
- error correction term 487
- estimation 495
- lag order specification 492
- retrieving coefficients 496
- trend assumption 490

Vector autoregression 477

Vector error correction 487

Verbose mode 581

Vertical line in graph 220

View 42

- default 44

Volatility 360

**W**

Wald test 330

- formula 330
- F-statistic 331
- joint restriction 333
- nonlinear restriction 334
- structural change with unequal variance 350

Warn on close option 580

Weibull distribution 575

Weighted least squares 251

- cross-equation weighting 456

nonlinear 264

nonlinear two stage 266

pool 539

system estimation 469

two stage 256

two stage in systems 457, 471

Weighting matrix

cross-section 548

for GMM 268, 473

heteroskedasticity and autocorrelation consistent 473

kernel options 473

White heteroskedasticity consistent 473

White

heteroskedasticity consistent covariance 251, 540, 549

test with no cross terms 340

Width of table column 225

Wilcoxon signed ranks test 150

Wildcard characters 39, 587

in easy query 120

Window

active 7, 45

database 103

EViews 10

object 47

Window control 8

Word processor 224, 225

Work area 13

Workfile

automatic backup 580

basics 33

create 35

directory 36

frequency 34, 583

range 34

resize 38

sample 58

save 36

sort 38

window 35

Write 533

## **X**

X-11 162

    limitations 162

XY-line

    graph 217

    view 182, 185

## **Y**

Yates' continuity correction 154

## **Z**

Zero line in graph 217