Notes in Macroeconometrics

Introduction to VAR Analysis

$$x = \begin{bmatrix} \Theta \\ p \\ m \\ r \end{bmatrix}$$

Jarir Ajluni

June 2005

I. Intro

Vector Autoregressions **VARs** were introduced initially as a replacement of large scale macroeconometric models estimated usually by OLS and GMM (Instrumental Variable regression), consisting of a huge set of equations estimated separately then parameters are subtracted in other equations. The VAR approach relies on a decreased amount of incredible restrictions and is a multi-equation regression system where uncertainty on which variables are actually exogenous provides the motivation.

In a VAR, consisting of two variables X,Y the path of Y is explained by the current and past realisations of X (and other variables additively) simultaneously the realisations of X rely on past and current realisations of Y, VARs introduction in the literature was done by Sims (1980) in an influential paper '*Macroeconomics & Reality*'.

Motivations for VAR

VARs were initially introduced and used for better forecasting performance of empirical econometric models, the main motivations for using VARs are listed below:

- VAR is a methodology adopted in replacement of large-scale macroeconometric models.
- VARs were found to improve forecasts.
- Based on the rejection of describing and determining exogenous variables as shown in the Lucas critique, Lucas (1976).
- VARs are made to ensure that lag lengths are appropriate and sufficient enough to overcome serial correlations, auto-correlated residuals in the presence of lagged dependent variables lead to Biased estimation.
- VARs overcome the unreasonable assumption of 'No feedback' VARs would ensure lack of feedback relations.

Context of VAR modelling

Multiple interests are involved in modelling VARs empirically, according to Favero (2001) the main context of VAR modelling are:

[1] Identify Shocks (particularly monetary policy shocks).

[2] Impulse Response Functions.

[3] Perform experiment in a theoretical based model.

[4] Perform a comparison between two sets of responses.

II. Specification and Development of VARs

The path of Y is explained by the current and past realisations of X and
the path of X is explained by the current and past realisations of Y yields the following equations:

$$y_t = a_1 + \beta_{12} x_t + \gamma_{11} y_{t-1} + \gamma_{12} x_{t-1} + e_{yt} \quad .... \ (1)$$

$$x_t = a_2 + \beta_{21} y_t + \gamma_{21} y_{t-1} + \gamma_{22} x_{t-1} + e_{xt} \quad .... \ (2)$$

Equations (1)(2) are both individually an ARDL models but the system of (1) and (2) as a multi-equation system is a VAR of order one in its 'structural form', the order refers to the number of considered in the VAR, thus VARs could include k variables and lags of p.

for very obvious reasons the models of (1) and (2) are unviable for econometric estimation by standard estimation techniques, because of the problems of Multicollinearity, Heteroscedasticity and Autocorrelation, since at least both e_{yt}, e_{xt} are correlated with their own lags and are correlated with variables X, Y in which the contrary is assumed for standard econometric estimation.
Rearranging (1) and (2) yields:

$$y_t - \beta_{12}\, x_t = a_1 + \gamma_{11}\, y_{t-1} + \gamma_{12}\, x_{t-1} + e_{yt} \quad .... (1.1)$$

$$x_t - \beta_{21}\, y_t = a_2 + \gamma_{21}\, y_{t-1} + \gamma_{22}\, x_{t-1} + e_{xt} \quad .... (2.1)$$

The compact form would be:

$$\begin{bmatrix} 1 & -\beta_{12} \\ -\beta_{21} & 1 \end{bmatrix} \begin{bmatrix} y_t \\ x_t \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} + \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix} \begin{bmatrix} y_{t-1} \\ x_{t-1} \end{bmatrix} + \begin{bmatrix} e_{yt} \\ e_{xt} \end{bmatrix}$$

Define:
the vectors: $\mathrm{Z}'_t = [y_t \quad x_t]'$, $A'_{0t} = [a_1 \quad a_2]'$, $\upsilon'_t = [e_{yt} \quad e_{xt}]'$

vector Z represents the joint determined or **‘Endogenous’** variables, while vector A consists of two intercepts and vector υ is a vector of residuals.
define the matrices of:

$$\mathrm{B} = \begin{bmatrix} 1 & -\beta_{12} \\ -\beta_{21} & 1 \end{bmatrix} \quad \& \quad \mathrm{H} = \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix}$$

matrix B is called the matrix of **‘feedback coefficients’** and matrix H is the **‘Autoregressive Coefficients’** matrix
then the compact form above would be written more compactly and equivalently as:

$$\mathrm{B}\,\mathrm{Z}_t = A_0 + \mathrm{H}\,\mathrm{Z}_{t-1} + \upsilon_t \quad .... (3)$$

equation (3) is the VAR is its structural form, simplifying for Z:

$$\mathrm{Z}_t = \mathrm{B}^{-1} A_0 + \mathrm{B}^{-1}\,\mathrm{H}\;\mathrm{Z}_{t-1} + \mathrm{B}^{-1}\;\upsilon_t \quad .... (3.1)$$

the VAR is its **‘reduced form’** is given by:

$$\mathrm{Z}_t = \mathrm{A}_0 + \Phi_1\;\mathrm{Z}_{t-1} + \varepsilon_t \quad .... (4)$$

where:

$$\mathrm{A}_{0t} = \mathrm{B}^{-1} A_0 \;, \quad \Phi_1 = \mathrm{B}^{-1}\,\mathrm{H}, \quad \varepsilon_t = \mathrm{B}^{-1}\,\upsilon_t$$

equation (4) is the VAR in the **‘reduced form’**, Z is a $k \times 1$ vector of k endogenous variables included in the VAR, A_0 a $k \times 1$ vector of intercepts, Φ_i is a $k \times k$ matrix of autoregressive coefficients, and ε_t is a vector of residuals with the dimension of $k \times 1$.

Estimation of the VAR

Equations (4) and (5) could be re-written to define elements of vector Z and matrices A_0, Φ_i then the elements α_{i0} as the i[th] element of matrix A_0 and α_{ij} as an element in row i and column j of matrix Φ, hence the system of (4) could be expressed as:

$$y_t = \alpha_1 + \alpha_{11}\, y_{t-1} + \alpha_{12}\, x_{t-1} + \nu_{1t} \quad .... \ (4.1)$$

$$x_t = \alpha_2 + \alpha_{21}\, y_{t-1} + \alpha_{22}\, x_{t-1} + \nu_{2t} \quad .... \ (4.2)$$

the above (4.1) and (4.2) are the **standard form** of the VAR, each equation could be estimated by standard OLS, the order of the VAR is large enough wipe away serial correlation in the residuals and moreover, with the appropriate lag order OLS estimates of the α parameters are consistent and asymptotically efficient. Those parameters are the estimates obtained once VAR is fitted using software packages as STATA 8.0, Microfit and E-views.

In the presence of an exogenous variable and a trend, the equations in (4.1)-(4.2) would include an estimated parameter for W and for the time trend t in each of the equations.

The Augmented VAR

The above is a VAR of order one, the VAR could be augmented in a generalised form to include a ρ number of lags and further more a time trend and an exogenous variable/vector could be included, the discussion of exogenous variables is also discussed below.

An Augmented $VAR(\rho)$ is given by:

$$Z_t = A_0 + \sum_{i=1}^{\rho} \Phi_i\, Z_{t-i} + \delta t + \Psi_t\, W_t + \varepsilon_t \ .... \ (5)$$

where: t is a time trend of 1, 2, 3, 4, 5 ….. T

W : a $k \times 1$ vector of k exogenous variables.

III. Structural VAR & Identification

In the structural VAR outlined in equations (1) & (2), the VAR is not viable for econometric estimation using standard regression techniques which assume that the explanatory variables are not correlated with the error terms e_{yt}, e_{xt}.

The only way for identifying a VAR is achievable only if we are willing to restrict the primitive system, so called structural system, a restricted VAR model is called a Structural VAR (SVAR) whereas there exist some restrictions on the parameters of the primitive system, either on the feedback coefficients (matrix B) or on the autoregressive coefficients (matrix H) the reason is clarified by the number of parameter estimated in the VAR and in the Structural SVAR, the meaning of restrictions that coefficients could be restricted to equal zero or to any other value very often suggested by an economic doctrine, incorporating a meaningful economic theory into the VAR modelling suggest our restrictions and the SVAR is then specified according to a theory-based restrictions instead of the unrestricted atheoretical VAR, in the SVAR some or all of the feedback and/or autoregressive coefficients are restricted, leaving always a number of k more parameters (where k denotes variables in the VAR), a VAR with k variables there is a number of $(k \times k) - k$ feedback coefficients and a number of $k \times k$ Auto-regressive coefficients.

A model with (k) variables is "exactly identified" if only $(k \times k) - k$ number coefficients are restricted, if more than $(k \times k) - k$ are restricted then the model is 'overidentified', consequently we can conclude that the VAR in its standard form of equations 4.1-2 is Underidentified.

Identification Schemes

As mentioned above imposing restrictions on either one or both of the feedback, autoregressive matrices, defines the SVAR, going through the process of the derivation of the standard form as in 4.1-2 it is obvious that the system would be different even if there is only one parameter restricted, the identifications schemes however varies, one type is the *Recursive identification*; imposing restrictions solely on the autoregressive coefficients also introduced as block granger causality restrictions (lags of a variable does not explain another variable in the system), or the restrictions could be contemporaneous, restricting the feedback matrix, so far the ordering of the variables in the vector was not important but to for identification the ordering become vital as to ensure exact identification it is a standard procedure to impose a lower triangular Choleski matrix decomposition with a number of restrictions equal to $(k \times k) - k$ for a matrix of the dimension $k \times k$.

Applying on the VAR above, assuming that there is no contemporaneous effect of variable *x* on *y* (then $\beta_{12} = 0$) and if we block granger causality between the two variables, the feedback and the autoregressive matrices B, H are restricted as shown below in the new matrices $\tilde{\mathrm{B}}, \tilde{\mathrm{H}}$:

$$\tilde{\mathrm{B}} = \begin{bmatrix} 1 & 0 \\ -\beta_{21} & 1 \end{bmatrix} \quad \tilde{\mathrm{H}} = \begin{bmatrix} \gamma_{11} & 0 \\ 0 & \gamma_{22} \end{bmatrix}$$

Matrix $\tilde{\mathrm{B}}$ is an example of an identification based on Choleski lower triangular decomposition, when Choleski type of identification is applied, ordering of the variables is crucial, in matrix $\tilde{\mathrm{H}}$ restrictions are an example of block-granger-causality type, the SVAR would be expressed as:

$$\begin{bmatrix} 1 & 0 \\ -\beta_{21} & 1 \end{bmatrix} \begin{bmatrix} y_t \\ x_t \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} + \begin{bmatrix} \gamma_{11} & 0 \\ 0 & \gamma_{22} \end{bmatrix} \begin{bmatrix} y_{t-1} \\ x_{t-1} \end{bmatrix} + \begin{bmatrix} e_{yt} \\ e_{xt} \end{bmatrix}$$

We then follow the same steps to obtain the standard form for estimation as in the equations from (3) till equations (4.1 – 4.2).

Short-run and Long run Identification

Restrictions on both the feedback and/or the autoregressive matrices are used to identify short run dynamics of a model, furthermore identification is obtained by imposing restrictions on the vector of disturbances, motivated by theoretical justifications that a shock in a given variable is does not effect some other variable(s).
Any SVAR model with no exogenous variables is written as:

$$\mathrm{A}\mathrm{Z}_t = \mathrm{A}_0 + \sum_{i=1}^{\rho} \Phi_i \, \mathrm{Z}_{t-i} + \mathrm{B}\varepsilon_t \; .... (6)$$

Where A, B are restrictions matrices of dimension equal to $(k \times k)$, now define the Lag operator:

$$A_i \, L^i X_t = A_i X_{t-p} \quad , i = 1, ... p$$

Then by applying the lag operator and taking the Z vector as common factor a VAR could be identically re-written in the form of:

$$\mathrm{Z}_t = \mathrm{A}_0 + A_1 L \, \mathrm{Z}_t + A_2 L^2 \, \mathrm{Z}_t + ... A_p L^p \, \mathrm{Z}_t + \varepsilon_t$$
$$\left(\mathrm{I}_K - \mathrm{A}_0 - A_1 L - A_2 L^2 - .... A_p L^p\right)\mathrm{Z}_t = \varepsilon_t \quad ....(7)$$

Hence the SVAR equivalent to (6) is given by:

$$\mathrm{A}\left(\mathrm{I}_K - \mathrm{A}_0 - A_1 L - A_2 L^2 - .... A_p L^p\right)\mathrm{Z}_t = \mathrm{A}\varepsilon_t \quad ....(8)$$

Where A, B are restrictions matrices of dimension equal to $(k \times k)$, Z is a vector of dimension of $(k \times 1)$, I_K is identity matrix $\left(A_{K \times K} / A_{K \times K}\right)$, A_0 is a $(k \times 1)$ vector of intercepts, $A_1 ....A_P$ are each $(k \times k)$ matrices of parameters, the model (8) is still not equivalent to (6) as matrix A multiplied by the vector of disturbances is not the orthogonalised matrix of B, vector ε_t is -subject to the proper lag lengths- is a normally distributed and serially uncorrelated satisfying the condition $\lfloor \varepsilon_t \varepsilon'_j \rfloor = 0_K \;, t \neq j$, define the $(k \times 1)$ vector of the orthogonalised disturbances v distributed normally with orthogonal variance $v \sim N(0, \mathrm{I}_K)$ also satisfying $\lfloor v_t v'_j \rfloor = 0_K \;, t \neq j$, hence the equality is:

$$\mathrm{A}\varepsilon_t = \mathrm{B}v_t$$
$$\mathrm{A}\varepsilon_t\varepsilon'_t\mathrm{A}' = \mathrm{B}v_t v'_t\mathrm{B}'$$

This is the proper definition of the matrix B, all of the above is a short run SVAR where clearly matrix A models the contemporaneous relations among the variables in the vector *Z*, and matrix B models the innovation (shocks) in each variable and its effects on the other variables to be restricted by a theory or the framework, usually matrix B is identified diagonally with zeros off the diagonal.

In order to capture the long run responses to both variables to their own shocks and to shocks in the other variables in the system, we need to identify the shocks of the vector of the disturbances, and set the feedback matrix to be an Identity matrix, recall the short run SVAR in (8) incorporating the above equality it becomes:

$$\mathrm{A}\left(\mathrm{I}_K - \mathrm{A}_0 - A_1 L - A_2 L^2 - \dots\dots\dots\dots A_p L^p\right)\mathrm{Z}_t = \mathrm{B}\,v_t \quad \dots\dots\dots(9)$$

setting $\mathrm{A}_{K\times K} = \mathrm{I}_K$ and define $L = \left(\mathrm{I}_K - \mathrm{A}_0 - A_1 L - A_2 L^2 - \dots\dots\dots\dots A_p L^p\right)$, expression in (9) is written compactly as:

$$\mathrm{Z}_t = L^{-1}\mathrm{B}\varepsilon_t$$

Now define the matrix of long run restrictions reflecting the effects of orthogonalised shocks as $C = L^{-1}\mathrm{B}$, then the long run SVAR is very compactly written in equation (10) below:

$$\mathrm{Z}_t = C\,v_t \quad \dots\dots\dots\dots(10)$$

Long run constraints are placed on the matrix C, matrices A, B and C are all assumed to be non-singular[1]

IV. Innovation Accounting

Once the VAR is fitted, an atheoretical VAR or an identified VAR on its own is not solely sufficient for a meaningful interpretation a very important applications are worthwhile, the Impulse Response Functions IRFs and the Forecast Error Variance Decomposition FEVDs.

The Impulse Response Function

The VAR could be also be expressed as a Vector Moving Average (VMA), from the VAR in (4) a VMA representation would take the form given below:

$$\mathrm{Z}_t = u + \sum_{i=0}^{\infty} \Phi_i \; \varepsilon_{t-i} \quad \dots\dots\dots \quad (11)$$

Where $u = [\bar{x} \;\; \bar{y}]'$, Φ_i is the matrix of the Impulse response coefficients, (8) is decomposed into:

$$\mathrm{Z}_t = \begin{bmatrix} \bar{y} \\ \bar{x} \end{bmatrix} + \sum_{i=0}^{\infty} \begin{bmatrix} \phi_{11}(i) & \phi_{12}(i) \\ \phi_{21}(i) & \phi_{22}(i) \end{bmatrix} \begin{bmatrix} \varepsilon_{y\,t-i} \\ \varepsilon_{x\,t-i} \end{bmatrix} \quad \dots\dots\dots\dots (11.1)$$

[1] Naming of the matrices A,B and C is consistent with STATA 8.2 codes and labelling of the estimation output.

Each of the $\phi_{ij}(i)$ is called the Impulse Response Function relating to the time period denoted by i, $\phi_{11}(0)$ for example is the instantaneous effect of a shock in the residuals of y on x, also known as the ***Impact Multiplier***.
The infinity term ∞ is replaced by an assigned horizon H then the whole effects could be accumulated in the form of:

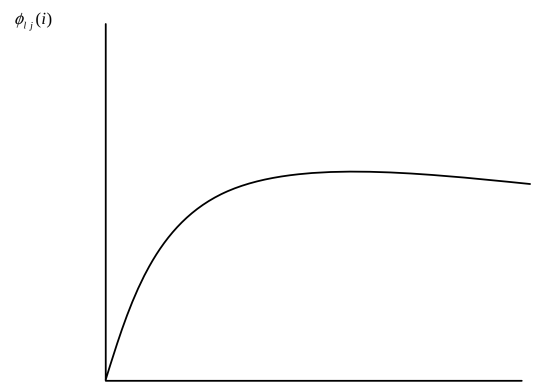
$\sum_{i=0}^{H} \phi_{lj}(i)$, for quarterly data usually the horizon ranges from 20 to 50 quarters ahead.

The IRF are very useful in the following aspects:

1. Identify the Dynamics of a model.
2. Determine over time the evolution of the effects of shocks on variables in the model, and
3. Measurement of the persistence of effects to these shocks.

IRFs are usually used in the form of a graph rather than numerically, by plotting ϕ_{lj} on (i)
An example below:

The IRF show the effect of shocks by long-run multiplier, showing the effects on y, x in response to various shocks, if all parameters in the primitive system are known, then its possible to trace the paths of the effects of a pure shocks in e_{yt}, e_{xt}, this methodology is not available as long as the VAR is under-identified, in order to identify the IRF the VAR should be restricted.



Choleski lower triangular matrix decomposition is a popular way to identify a VAR system as the Choleski identification provides minimal set of assumptions that can be used to identify the structural model, and here the ordering of the variables within the VAR becomes crucial, the importance of the ordering depends on the magnitude of the correlation between the shocks In the residual vector of the VAR, the practical empirical suggestion is to list the variables with *the most endogenous variable ordered last*

Forecast Error Variance Decomposition

Is an innovation accounting of the effects of a variable i on variable j measuring the fraction of the error in forecasting j all along a specified horizon h in which is attributable to innovations in variable j.

The forecast error from the VAR in (4) of vector Z where $Z'_t = [y_t \quad x_t]'$, from the VAR in (4) it is clear that the value of the forecasted Z_{t+1} is given by the following VAR:

$$Z_{t+1} = A_0 + \Phi_1 Z_t + \varepsilon_{t+1}$$

the one-step forecast error is then given by:

$$Z_{t+1} - E_t Z_{t+1} = \varepsilon_{t+1}$$

from the above expression, the VAR suggests that $E_t\, Z_{t+1} = A_0 + \Phi_1\, Z_t$, which makes the expected value of Z the predicted value by the VAR denoted by $\hat{Z}_{t+1}$, then the forecast error over a given horizon h, FE^h is as follows:

$$Z_{t+h} - \hat{Z}_{t+h} = \sum_{i=0}^{h-1} \Phi_i\, \varepsilon_{t+h-i} \quad .... (12)$$

where

Z_{t+h}, is the actual observed values, $\hat{Z}_{t+h}$ is the h-step predicted values estimated at time t.

Decomposing Z into its components, using $Z'_t = [y_t \quad x_t]'$ and decomposing the matrix of Φ_i into its elements of $\phi_{ab}(h)$ derives the FEVD by showing the forecast error for y solely in the following form:

$$y_{t+h} - \hat{y}_{t+h} = \phi_{11}(0)\varepsilon_{y\,t+1} + \phi_{11}(1)\varepsilon_{y\,t+2} + + ... + \phi_{11}(h-1)\varepsilon_{y\,t+h} +$$
$$\phi_{12}(0)\varepsilon_{x\,t+1} + \phi_{12}(1)\varepsilon_{x\,t+1} + ... + \phi_{12}(h)\varepsilon_{x\,t+1} \qquad ....(13)$$

the next step is the vital importance of expressing the squared forecast error as a variance, since the definition of a variance is $\sigma_X^2 = X_t - E_t X_t$ for any variable *X*, then $(y_{t+h} - E_t y_{t+h})^2 = \sigma_y^2(h)$, since $(y_{t+h} - E_t y_{t+h})$ is identical to $(y_{t+h} - \hat{y}_{t+h})$, then it is clear that B could be given in the Forecast Error Variance form:

$$\sigma_y^2(h) = \phi_{11}^2(0)\varepsilon_{y\,t+1} + \phi_{11}^2(1)\varepsilon_{y\,t+2} + + ... + \phi_{11}^2(h-1)\varepsilon_{y\,t+h} +$$
$$\phi_{12}^2(0)\varepsilon_{x\,t+1} + \phi_{12}^2(1)\varepsilon_{x\,t+1} + ... + \phi_{12}^2(h)\varepsilon_{x\,t+1} \qquad ....(13)$$

This expression show the forecast error variance of y that is attributable to both y and x and using this expression a fraction of this variance is attributable only to y itself by the shocks of ε_{yt} and the rest is attributable on x by the shocks of ε_{xt}.

$$\sigma_y^2\left[\phi_{11}^2(0)\varepsilon_{y\,t+1} + \phi_{11}^2(1)\varepsilon_{y\,t+2} + + ... + \phi_{11}^2(h-1)\varepsilon_{y\,t+h}\right]/\sigma_y^2(h) \quad ....(13.1)$$
$$\sigma_x^2\left[\phi_{12}^2(0)\varepsilon_{x\,t+1} + \phi_{12}^2(1)\varepsilon_{x\,t+2} + + ... + \phi_{12}^2(h-1)\varepsilon_{x\,t+h}\right]/\sigma_x^2(h) \quad ....(13.2)$$

(13.1) is the FEVD of series y that is attributable to its own shocks, while (13.2) is the variance decomposition of series y attributable to shocks in x.

The usefulness of the FEVD is that the unrestricted VAR is obviously over-parameterised for short horizon forecasting, more important is the information FEVD provides on the proportion of the movement of a variable to be either due to its own shocks or to other variable(s) shocks, this is important as effects on a variable provides evidence on exogeneity or endogeneity, if the value of (13.1) is very high and (13.2) is very low in the above expression this means that a high proportion of the forecast error in y is determined by its own shocks and hence it is concluded that y is exogenous.

On finding empirical results on the relationships between variables and more specifically on the effects of shocks of some indicating variables (as interest rates measuring monetary policy shocks) the FEVD is a very useful tool of identifying the impact of shocks with interpretable findings.

V. Post estimation VAR tests:

Once VAR is fitted there are several statistical tests of interest to a relevant aspects of the empirical application of the VAR on actual data, the normality of the error terms of the residual vector is of central importance to interpret VAR results, to achieve this normality the lag length should be long enough, furthermore residual Autocorrelation just as in OLS is also an important feature.

[1] Wald lag exclusion test statistic:

This test the hypothesis that the endogenous variables at a given lag are jointly indifferent from zero, Wald test is applied on either one or all equations.
This test is often applied on an unrestricted VAR in order to motivate a recursive identification scheme (restrictions on some of the lagged endogenous variables).

[2] Lagrange Multiplier Statistic for residual serial correlation

The assumption of non-serially-correlated errors in a VAR is rather vague, the unsatisfactory assumption is tested by a test introduced in Johansen (1995), that presented a serial autocorrelation test in a VAR context with the null of No Autocorrelation, the alternative imply presence of autocorrelation and hence lag length should be altered, the model is also suitable for testing Autocorrelation at several lag lengths.

The procedure is once a VAR is fitted then the residuals are estimated and saved, these residual vectors of $v_1, v_2, v_3, ...v_k$ (from 4.1-4.2) then treating these residuals as variables the original VAR is augmented with lags of these residuals as variables at a given lag length; an order of ρ then the residuals are lagged ρ times.
The LM statistic at lag j is:

$$LM_{\rho} = (T - d - 0.5)\ \ln\left(\frac{\left|\hat{\Sigma}\right|}{\left|\tilde{\Sigma}_{\rho}\right|}\right) \quad .... \quad (14)$$

T: number of observations.
$\hat{\Sigma}$: Maximum likelihood estimates of the variance covariance matrix Σ.
$\tilde{\Sigma}_{\rho}$: Maximum likelihood estimates of the variance covariance matrix Σ from the augmented regression on the lagged estimated and saved residuals.
$LM_{\rho} \sim \chi^2 (k^2)$

The LM statistic has a chi-square distribution with k^2 degrees of freedom; the test is implemented in comparing the test statistic with the tabulated values under the null hypothesis of No Autocorrelation and with the alternative of Autocorrelation.

[3] Testing Normality

A series of statistics could be constructed against the null that residuals in a VAR are normally distributed, these tests could be applied to either one or all equations of the system, the tests include a test of ***Skewness***, the second test is for ***Kurtosis*** and lastly both could be combined in the Jarque-Bera test statistic.

Skewness: is a measurement of the asymmetry of a distribution, the bell shape of the normal distribution is asymmetrical in the presence of an excessive Skewness (meaning one tail is longer than the other).
Skewness is given by the equation:

$$Skewness = \frac{\mu_3}{\sigma^3} \;, \quad \mu_3 = (x - \bar{x})^3$$

Kurtosis: measurement of the "peakedness"; where the peak of the normal distribution with the bell shape is stretched up or pressured down, a normal distribution should be *mesokurtic* (with zero Kurtosis), the meaning of Kurtosis is that the variance is due to infrequent deviations or shocks, in the presence of Kurtosis a distribution is called leptokurtic (higher peak and fat tails) or platykurtic (smaller peak and thinner tails).
Kurtosis is given by the equation:

$$\gamma = \frac{\mu_4}{\sigma^4} - 3 \;, \quad \mu_4 = (x - \bar{x})^4$$

The above term is known as the Kurtosis excess with a –3 to adjust to the normal distribution, the Jarque-Bera statistic would test both Skewness & Kurtosis, if the Jarque-Bera statistic is equal to zero then Skewness is 0 and kurtosis is 3 then the distribution is normal, and hence we statistically test the normality.
The Jarque-Bera test is given by:

$$JB = \frac{N}{6}\left[S^2 + \frac{(K-3)^2}{4}\right] \approx \chi_2^2 \quad \text{.............} (15)$$

The above expression applies to each residual in the VAR estimated equations, the JB statistic has a Chi-square distribution with 2 degrees of freedom per equation compared with the Chi-square probability tables to reject the null of normality, hence a VAR of four endogenous variables would have 8 degrees of freedom.

VI. VAR In Modelling Monetary Policy: Literature Review of 'Puzzles'

The VAR approach for modelling monetary transmission mechanism has been criticised on the basis that it views Central Banks as 'random numbers generators' rather than an institutions regulating the economy, it views the economy as a machine which is incorrect, monetary policy when estimated using a SVAR the focus is not on rules but on deviations from the rules, or on shocks that affect the system, "*since only when Central Banks deviate from their rules it becomes possible to collect interesting information on the response of macroeconomic variables to monetary policy impulses*"[2].

Imposing restrictions is a necessary prerequisites for the analysis, shocks are identified by Choleski decomposition viable for econometric estimation unlike the SVAR, where there is a crucial assumption that structural shocks are linear combinations of the residuals in the reduced form as in equation (4), Lippi and Reichlin (1993) argue that modern macroeconomic models which are "linearlised" into dynamic system tend to have a non-invertible moving average components, and structural shocks are therefore unidentifiable.

Sims Interpretation of Monetary Policy Shocks: Sims (1992)

Sims (1992) suggested that innovations in the monetary aggregates might not correctly represent changes in monetary policy, the monetary policy is "operated" by setting the interest rates, reflect the actual monetary policy, Central Banks announce interest rates in a monthly or quarterly basis, as Sims shows, the innovations of interest rates are often due to inflationary pressures in the economy.

Other interpretations suggest using narrower monetary aggregates as in the Eichenbaum and Evans (1995) use of the non-borrowed reserves, Sims' interpretation is very often adopted in numerous empirical studies and in identifications of VAR systems in the literature.

[2] Favero (2004), p. 173

Kim & Roubini (2000) JME

The likely effects of monetary contraction

Kim and Roubini (2000) show expected movements of each macroeconomic variable by a monetary contraction initially and in the longer horizon as follows:

- Interest Rates: initially interest rates should rise.
- Monetary aggregates: such as M2 should initially fall hence it's a monetary contraction.
- Price level: declines as suggested by most theoretical models.
- Output: remains constant after monetary policy contraction.
- Exchange rate: depends on the expected inflation, a rise of interest rates is usually related with appreciation of exchange rate, KR argue that if the expected inflation is high then a rise in interest rates (monetary contraction) is associated with exchange rate depreciation. On the exchange rate there is several theories suggesting conflicting responses to monetary policy shocks, the response on the Exchange rate would depend on whether the nominal or real interest rate that is increasing, the analysis could adopt either the flexible price models or the flexible price with money neutrality, or adopting the sticky price overshooting model of Dornbusch (1976), both overshooting and flexible price models suggest a positive monetary innovation will devalue the domestic currency, exchange rate depreciation, also using the UIP then an increase in the domestic interest rate (monetary contraction) should depreciate the exchange rate.

Monetary Policy Puzzles

The vast majority of empirical analysis on the effects of monetary policy has been plagued by a number of puzzles in the responses of the all-macroeconomic variables:

- The Liquidity Puzzle:

  Identifying monetary policy as innovations in monetary aggregates, such innovations appear to be associated with an increase and not a decrease in the exchange rate, no devaluation is a very unexpected and unjustified by any theory.

- The Price Puzzle:

  Identifying monetary policy as innovations in the interest rates, the response of prices are "wrong" as the monetary tightening is associated with an increase rather than a decrease in the price level, empirical results show a reduction of interest rates increasing money is leading to an increase in prices, making an unexplained puzzle.

- Exchange rate Puzzle

  Text book models of the 1960s and 1970s easily simply follow this sequence; raising the interest rates attract foreign funds and this increase demand for domestic currency leading to an appreciation of the exchange rate, the argument ends with no consideration of alternative responses and to other things not being constant, in empirical results unlike the case for the United States a positive innovation in the interest rates depreciated the exchange rates of the economies in the rest of G7, as was found in a study by Grilli and Roubini (1995) and also in Sims (1992), while In the US interest rates increases lead to appreciation of the dollar.

- The Forward Discount Puzzle

  If the UIP holds then it is obvious that the exchange rate should depreciate if there is an increase in the domestic interest rates, empirical evidence suggested a persistent appreciation of the exchange rate, ruling out entirely the UIP and forming a forward discount puzzle.

VII. Concluding Remarks: Criticism of the VAR approach

VARs are criticised of being atheoretical -despite improving forecasts- it's use is questionable as it leads to a difficult tasks of interpretation and identification of relationships, the SVAR imposes theoretical priori restrictions providing an interesting tool for examining theoretical frameworks, from econometric point of view over-parameterization is the main shortcomings of a VAR, another problem from both econometric and economic perspectives is the importance of having an independently identically distributed vector of residuals, in which most of the VAR practices would become misleading if it exhibits excessive Skewness or kurtosis, or both.
Also the assumption of structural shocks being linear combinations of the residuals which is argued by Lippi and Reichlin (1993) as explained earlier on the basis that "linearlisation" would lead to having a non-invertible moving average components, and therefore structural shocks are unidentifiable. In a very recent paper Chari, Kehoe and McGrattan (2005) argue that VARs with long run restrictions could be very misleading given "quantitatively" false inference and SVARs are still incorporating little economic theory, according to Chari, Kehoe and McGrattan (2005) SVARs approach is still "an illustrative or corroborative method".

References

[1] Blanchard, O.J and Quah, D, (1989), "The Dynamic Effects of Aggregate Demand and Supply Disturbances", *American Economic Review*, 79, 655-673.

[2] Chari, V.V., Kehoe P. and McGrattan, E. (2005), 'A critique of structural VARs using real business cycle theory', *Federal Reserve Bank of Minneapolis* working paper 631, February.

[3] Dornbusch, Rudiger (1976), "Expectations and Exchange Rate Dynamics", *Journal of Political Economy*, 84(6), p. 1161-76.

[4] Eichenbaum, M and C.L. Evans (1995), "Some Empirical Evidence on the Effects of Shocks to Monetary Policy on Exchange Rates," *Quarterly Journal of Economics*, 110, pp. 975-1009 rd. Amsterdam: North-Holland, Elsevier.

[5] Enders, Walter, (2004) "Applied Time Series Econometrics", 2nd Edition, Wiley publishing.

[6] Favero, Carlo (2001), "Applied Macroeconometrics", Oxford University Press.

[7] Grilli, V. and Roubini, N (1995) "Liquidity and Exchange Rates: Puzzling evidence from the G-7 Countries", Mimeo, Yale University.

[8] Johansen, S., (1995) "Likelihood-Based Inference in Cointegrated Vector Autoregressive models", Oxford, UK: Oxford University Press.

[9] Lippi, M and Reichlin, L, (1993), "Dynamic Effects of Aggregate Demand and Supply Disturbances: comment" *American Economic Review*, 83, 644-652.

[10] Lucas, (1976), " Economic Policy: A Critique", *Carnegie-Rochester Conference Series on Public Policy*, 1, 19-46.

[11] Kim, S. and Roubini, N, (2000), "Liquidity and Exchange Rates in the G7 Countries: Evidence from Identified Vars", *mimeo*

[12] Kim, S. and Roubini, N, (2000), "Exchange Rate Anomalies in the Industrial Countries: A Solution with a Structural VAR Approach", *Journal of Monetary Economics*, 45, pp.561-586.

[13] Sims, C, (1980), "Macroeconomics and Reality", *Econometrica*, 48, 1-48.

[14] Sims, C, (1992), "Interpreting the Macroeconomic Time Series Facts: The Effect of Monetary Policy", *European Economic Review*, 36, 975-1000.