MGMG 522 : Session #7

Serial Correlation

(Ch. 9)

Serial Correlation (a.k.a. Autocorrelation)

- Autocorrelation is a violation of the classical assumption #4 (error terms are uncorrelated with each other).
- Autocorrelation can be categorized into 2 kinds:
  - Pure autocorrelation (autocorrelation that exists in a correctly specified regression model).
  - Impure autocorrelation (autocorrelation that is caused by specification errors: omitted variables or incorrect functional form).
- Autocorrelation mostly happens in a data set where order of observations has some meaning (e.g. time-series data).

Pure Autocorrelation

- Classical assumption #4 implies that there is no correlation between any two observations of the error term, or $E(r_{ij}) = 0$ for $i \neq j$.
- Most common kind of autocorrelation is the <u>first-order</u> autocorrelation, where current observation of the error term is correlated with the previous observation of the error term.
- Mathematically, $\varepsilon_t = \rho\varepsilon_{t-1} + u_t$.
  Where, ε = error term, ρ = simple correlation coefficient ($-1 < \rho < +1$), and u = classical error term.

7-3

ρ (pronounced "rho")

- $-1 < \rho < +1$
- $\rho < 0$ indicates negative autocorrelation (the signs of the error term switch back and forth).
- $\rho > 0$ indicates positive autocorrelation (a positive error term tends to be followed by a positive error term and a negative error term tends to be followed by a negative error term).
- Positive autocorrelation is more common than negative autocorrelation.

7-4

Higher-order Autocorrelation

- Examples:

1. Seasonal autocorrelation:

   $\varepsilon_t = \rho\varepsilon_{t-4} + u_t$

2. Second-order autocorrelation:

   $\varepsilon_t = \rho_1\varepsilon_{t-1} + \rho_2\varepsilon_{t-2} + u_t$.

7-5

Impure Autocorrelation

- Caused by specification errors: omitted variables or incorrect functional form.
- Specification errors should be corrected first by way of investigating the independent variables and/or functional form.
- How can omitted variables or incorrect functional form cause autocorrelation?
- Remember that the error term is the sum of the effects of:
  1. Omitted variables
  2. Nonlinearities
  3. Measurement errors
  4. Pure error

7-6

Example: Omitted Variable Causes Autocorrelation

- Correct model: $Y = \beta_0+\beta_1X_1+\beta_2X_2+\varepsilon$
- If X_2 is omitted: $Y = \beta_0+\beta_1X_1+\varepsilon^*$
- Where, $\varepsilon^* = \beta_2X_2+\varepsilon$
- If ε is small compared to β_2X_2, and X_2 is serially correlated (very likely in a time series), ε^* will be autocorrelated.
- Estimate of β_1 will be biased (because X_2 is omitted).
- Both the bias and impure autocorrelation will disappear once the model gets corrected.

7-7

Example: Incorrect Functional Form Causes Autocorrelation

- Correct model: $Y = \beta_0+\beta_1X_1+\beta_2X_1^2+\varepsilon$
- Our model: $Y = \beta_0+\beta_1X_1+\varepsilon^*$
- Where, $\varepsilon^* = \beta_2X_1^2+\varepsilon$
- Autocorrelation could result. (See Figure 9.5 on p. 323)

7-8

Consequences of Autocorrelation

1. Pure autocorrelation does not cause bias in the coefficient estimates.
2. Pure autocorrelation increases variances of the coefficient estimates.
3. Pure autocorrelation causes OLS to underestimate the standard errors of the coefficients. (Hence, pure autocorrelation overestimates the t-values.)

7-9

Example of the Consequences

1. With no autocorrelation
   - b_1 = 0.008
   - SE(b_1) = 0.002
   - t-value = 4.00
2. With autocorrelation but a correct SE of coefficient
   - b_1 = 0.008
   - $SE^*(b_1)$ = 0.006
   - t-value = 1.33
3. With autocorrelation and OLS underestimate SE of coefficient
   - b_1 = 0.008
   - SE(b_1) = 0.003
   - t-value = 2.66

7-10

Detection of Autocorrelation

- Use the Durbin-Watson d Test
- Durbin-Watson d Test is only appropriate for
  - a regression model with an intercept term,
  - autocorrelation is of first-order, and
  - The regression model does not include a lagged dependent variable as an independent variable (e.g., Y_{t-1})
- Durbin-Watson d statistic for T observations is: $d = \frac{\sum_{2}^{T}(e_t - e_{t-1})^2}{\sum_{1}^{T} e_t^2}$

7-11

d statistic

- d = 0 indicates extreme positive autocorrelation ($e_t = e_{t-1}$).
- d = 4 indicates extreme negative autocorrelation ($e_t = -e_{t-1}$).
- d = 2 indicates no autocorrelation $\Sigma(e_t - e_{t-1})^2 = \Sigma(e_t^2 - 2e_t e_{t-1} + e_{t-1}^2) = \Sigma(e_t^2 + e_{t-1}^2)$.

7-12

The Use of d Test

1. Econometricians almost never test one-sided that there is negative autocorrelation. Most of the tests are to detect positive autocorrelation (one-sided) or to detect autocorrelation (two-sided).
2. d test is sometimes inconclusive.

7-13

Example: One-sided d test that there is positive autocorrelation

H_0: $\rho \leq 0$ (no positive autocorrelation)

H_1: $\rho > 0$ (positive autocorrelation)

- Decision Rule

If $d < d_L$	Reject H_0
If $d > d_U$	Do not reject H_0
If $d_L \leq d \leq d_U$	Inconclusive

7-14

Example: Two-sided d test that there is autocorrelation

H_0: $\rho = 0$ (no autocorrelation)
H_1: $\rho \neq 0$ (autocorrelation)

- Decision Rule

If $d < d_L$	Reject H_0
If $d > 4-d_L$	Reject H_0
If $4-d_U > d > d_U$	Do not reject H_0
Otherwise	Inconclusive

Correcting Autocorrelation

- Use the Generalized Least Squares to restore the minimum variance property of the OLS estimation.

$$Y_t = \beta_0 + \beta_1 X_{1t} + \varepsilon_t \rightarrow (\text{Eq. }1)$$

$$\varepsilon_t = \rho\varepsilon_{t-1} + u_t$$

$$Y_t = \beta_0 + \beta_1 X_{1t} + \rho\varepsilon_{t-1} + u_t \rightarrow (\text{Eq. }2)$$

$$\rho * (\text{Eq. }1) \text{ and lag 1 period: } \rho Y_{t-1} = \rho\beta_0 + \rho\beta_1 X_{1t-1} + \rho\varepsilon_{t-1} \rightarrow (\text{Eq. }3)$$

$$(\text{Eq. }2) - (\text{Eq. }3): \; Y_t - \rho Y_{t-1} = \beta_0(1-\rho) + \beta_1(X_{1t} - \rho X_{1t-1}) + u_t$$

$$Y_t^* = \beta_0^* + \beta_1 X_{1t}^* + u_t \rightarrow (\text{Eq. }4)$$

$$Y_t^* = Y_t - \rho Y_{t-1} \qquad X_{1t}^* = X_{1t} - \rho X_{1t-1} \qquad \beta_0^* = \beta_0 - \rho\beta_0$$

GLS properties

1. Now, the error term is not autocorrelated. Thus, OLS estimation of Eq.4 will be minimum variance.
2. The slope coefficient β_1 of Eq.1 will be the same as that of Eq.4, and has the same meaning.
3. Adj-R^2 from Eq.1 and Eq.4 should not be used for comparison because the dependent variables are not the same in the two models.

7-17

GLS methods

1. Use the Cochrane-Orcutt method (EViews does not support this estimation method.) (Details on p. 331-333)
2. Use the AR(1) method (In EViews, insert the term AR(1) after the list of independent variables.)
3. When d test is inconclusive, GLS should not be used.
4. When d test is conclusive, GLS should not be used if
   1. The autocorrelation is impure.
   2. The consequence of autocorrelation is minor.

7-18

Autocorrelation-Corrected Standard Errors

- The idea of this remedy goes like this.
- Since, there is no bias in the coefficient estimates.
- But, the standard errors of the coefficient estimates are larger with autocorrelation than without it.
- Therefore, why not fix the standard errors and leave the coefficient estimates alone?
- This method is referred to as HCCM (heteroskedasticity-consistent covariance matrix).

7-19

Autocorrelation-Corrected Standard Errors

- In EViews, you will choose “LS” and click on “Options”, then select “Heteroskedasticity-Consistent Coefficient Covariance” and select “Newey-West”.
- The standard errors of the coefficient estimates will be bigger than those from the OLS.
- Newey-West standard errors are also known as HAC standard errors; for they correct both Heteroskedasticity and AutoCorrelation problems.
- This method works best in a large sample data. In fact, when the sample size is large, you can always use HAC standard errors.

7-20