OSU Economics 444: Elementary Econometrics

Ch.9 Serial Correlation

- **Pure serial correlation** for time series data

\[ E(\epsilon_t\epsilon_s) \neq 0, \quad t \neq s. \]

1) *serial correlation:* If the expected value of two observations of the error term over time is not equal to zero, then the error term is said to be serially correlated.

2) *First-order serial correlation:*

\[ \epsilon_t = \rho \epsilon_{t-1} + u_t, \]

where
- \( \epsilon \) = the error term of the regression function,
- \( \rho \) = the parameter between the observations of the error term
- \( u \) = classical (nonserially correlated) error term.

(a) The process of \( \epsilon_t \) is a first-order autoregressive process (AR(1)). \( \rho \) is called the *first-order autocorrelation coefficient*.

(b) \( \rho \) measures the strength of the serial correlation. If \( \rho = 0 \), there is no serial correlation. If \( \rho \) approaches to one in absolute value, a high degree of serial correlation exists.

(c) The condition \( |\rho| < 1 \) ensures that \( \epsilon_t \) in AR(1) process has a finite variance. The variance of \( \epsilon_t \) is

\[ \sigma^2_\epsilon = \frac{\sigma^2_u}{(1-\rho^2)}, \]

where \( \sigma^2_u \) is the variance of \( u_t \).

(d) If \( \rho \) is equal to one, the process of \( \epsilon_t \) is called a *random walk*. If \( \rho \) is greater than one, the error term has a tendency to increase over time (explode).

(e) A positive \( \rho \) implies that the error term tends to have the same sign from one time period to the next. This is *positive serial correlation*. In economic time-series applications, positive correlation is the rule and not the exception.

Positive serial correlation Figure 9.1

no serial correlation Figure 9.2

3) *seasonally based serial correlation*

\[ \epsilon_t = \rho \epsilon_{t-4} + u_t. \]

In a quarterly model, the current quarter’s error term may be functionally related to the observation of the error term form in the same quarter of the previous year.

4) *Second-order serial correlation:*

\[ \epsilon_t = \rho_1 \epsilon_{t-1} + \rho_2 \epsilon_{t-2} + u_t. \]

- **Impure Serial Correlation**
— serial correlation caused by a specification error (such as omitted variable or an incorrect functional form).

1) omitted variable, e.g.,

true equation \( y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \epsilon_t \).

Then,

\( y_t = \beta_0 + \beta_1 x_{1t} + \epsilon^*_t \),

where \( \epsilon^*_t = \beta_2 x_{2t} + \epsilon_t \), is a misspecified equation.

The new error term \( \epsilon^* \) may tend to be serially correlated when \( x_2 \) itself is serially correlated and the variance of \( \epsilon_t \) is small compared to that of \( \beta_2 x_2 \).

2) incorrect functional form, e.g.,

true equation \( y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{1t}^2 + \epsilon_t \)

but the equation is misspecified as

\( y_t = \alpha_0 + \alpha_1 x_t + \epsilon^*_t \),

where \( \epsilon^*_t \) is now a function of the true error term \( \epsilon_t \) and of the differences between the linear and the polynomial functional forms.

Figure 9.5 incorrect functional form

3) The best remedy for impure serial correlation is to attempt to find the omitted variable or the correct functional form for the regression equation.

•• Consequences of (Pure) Serial Correlation

(1) The OLSE estimates are still centered around the true \( \beta \)s.

For example, consider \( y_t = \beta x_t + \epsilon_t \) with \( \epsilon_t = \rho \epsilon_{t-1} + u_t \). The OLSE of \( \beta \) is

\[
\hat{\beta} = \frac{\sum_{t=1}^{T} x_t y_t}{\sum_{t=1}^{T} x_t^2} = \beta + \frac{\sum_{t=1}^{T} x_t \epsilon_t}{\sum_{t=1}^{T} x_t^2},
\]

and, hence,

\[
E(\hat{\beta}) = \beta + \frac{\sum_{t=1}^{T} x_t E(\epsilon_t|x_1, \cdots, x_T)}{\sum_{t=1}^{T} x_t^2} = \beta,
\]

because \( E(\epsilon|x_1, \cdots, x_T) = 0 \).

(2) The Gauss-Markov theorem is no longer valid. So there is a possibility to get better estimates than the OLSEs.

(3) Serial correlation tends to increase the variance of the OLSE \( \hat{\beta} \). It can be shown that the variance of \( \text{OLSE} \hat{\beta} \) is a function of \( \rho \). The larger is the absolute value of \( \rho \), the larger the variance of the \( \hat{\beta} \) tends to be.
(a) For example, for the simple model \( y_t = \beta x_t + \epsilon_t \) with \( \epsilon_t = \rho \epsilon_{t-1} + \xi_t \), the (correct) variance of OLSE \( \hat{\beta} \) is

\[
\text{var}(\text{OLS} \, \hat{\beta}) = \frac{\sigma^2_u}{\sum_{t=1}^{T} x_t^2} \left( 1 + 2\rho \sum_{t=1}^{T-1} x_t x_{t+1} + 2\rho^2 \sum_{t=1}^{T-2} x_t x_{t+2} + \cdots + 2\rho^{T-1} \sum_{t=1}^{T} x_t^2 \right).
\]

In general, \( x_t \) would be positive correlated and \( \rho \) likely be positive. Then the variance of OLSE tends to increase as \( \rho \) increases.

(4) With serial correlation, the typical OLS formula for the standard error of OLSE does not apply. The formula yields values that tend to underestimate the proper standard deviation of the OLSE. In consequence, it causes the \( t \)-scores to be overestimated. It becomes more likely that it leads to over-reject the true hypothesis \( H_0 : \beta = 0 \).

(a) For the model \( y_t = \beta x_t + \epsilon_t \). The conventional (incorrect) formula on the computer for the OLS \( \hat{\beta} \) will estimate \( \frac{\sigma^2_u}{\sum_{t=1}^{T} x_t^2} \). But we know that

\[
\text{var}(\text{OLS} \, \hat{\beta}) \geq \frac{\sigma^2_u}{\sum_{t=1}^{T} x_t^2}.
\]

• The Durbin-Watson Statistic \( d \):

\[
d = \frac{\sum_{t=2}^{T}(e_t - e_{t-1})^2}{\sum_{t=1}^{T} e_t^2},
\]

where \( e_t \)'s are the OLS residuals.

1) The Durbin-Watson test is valid only if the regression equation includes an intercept term and it does not include a lagged dependent variable.

2) The Durbin-Watson test is designed to determine if there is any first-order serial correlation in the error term, i.e., \( e_t = \rho e_{t-1} + \xi_t \).

3) It has power also against higher order serial correlation.

• The possible values of \( d \) under various degrees of serial correlation (with \(-1 \leq \rho \leq 1\)):

1) When \( \rho = 1 \), i.e., extreme positive serial correlation, \( d \approx 0 \).

   \( \rho = 1 \) implies that \( e_t \approx e_{t-1} \) and, hence, \( d \approx 0 \).

2) When \( \rho = -1 \), i.e., extreme negative serial correlation, \( d \approx 4 \).

   \( \rho = -1 \) implies that \( e_t \approx -e_{t-1} \) and, hence,

\[
d \approx \frac{\sum_{t=2}^{T}(2e_t)^2}{\sum_{t=1}^{T} e_t^2} \approx 4.
\]

3) When \( \rho = 0 \), i.e., no serial correlation, \( d \approx 2 \).

\[
d = \frac{\sum_{t=2}^{T} e_t^2 - 2 \sum_{t=2}^{T} e_t e_{t-1} + \sum_{t=2}^{T} e_{t-1}^2}{\sum_{t=1}^{T} e_t^2} \approx \frac{\sum_{t=2}^{T} e_t^2 + \sum_{t=2}^{T} e_{t-1}^2}{\sum_{t=1}^{T} e_t^2} = 2,
\]

because \( \sum_{t=2}^{T} e_t e_{t-1} \approx 0 \) when \( \rho = 0 \).
• Using the Durbin-Watson Test

1) Econometrician has interest only in testing positive correlation (because economics time series usually does not have negative correlation).

2) The Durbin-Watson statistic has three regions – an acceptance region, a rejection region, and an inconclusive region.

3) The critical values \( d_L \) (lower bound) and \( d_U \) (upper bound) of the Durbin-Watson statistic can be found in published tables. They depend on sample size and the number of explanatory variables in the linear regression model.

4) Given \( H_0 \): no positive serial correlation and \( H_1 \): positive serial correlation, the decision rule is
   - 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.

5) An numerical example: 5 % level of significance, 3 explanatory variables (excluding constant term), 25 observations. The critical values are \( d_L = 1.12 \) and \( d_U = 1.66 \).
   a) Suppose \( d = 1.78 \), then there is no evidence of positive correlation.
   b) Suppose \( d = 1.28 \), then we can not make conclusion.
   c) Suppose \( d = 0.60 \), then there is positive correlation.

   (Figure 9.7 here)

• Generalized Least Squares – can be used to improve upon the OLSE when the disturbances are serially correlated.

1) GLS transforms the data (and the disturbances) such that the new disturbances meet all the assumptions of a classical linear regression model.

2) For the regression equation \( y_t = \beta_0 + \beta_1 x_t + \epsilon_t \) with \( \epsilon_t = \rho \epsilon_{t-1} + u_t, \ t = 1, \cdots, T \), one can transform them by a quasi-difference to

   \[
   (y_t - \rho y_{t-1}) = \beta_0 (1 - \rho) + \beta_1 (x_t - \rho x_{t-1}) + u_t, \quad t = 2, \cdots, T.
   \]

3) Let \( y_t^* = y_t - \rho y_{t-1}, \ x_t^* = x_t - \rho x_{t-1} \) and \( \beta_0^* = (1 - \rho) \beta_0 \). The quasi-differenced equation becomes

   \[
   y_t^* = \beta_0^* + \beta_1 x_t^* + u_t, \quad t = 2, \cdots, T. \tag{\star}
   \]

4) The quasi-differenced equation has used up \( y_1 \) in its difference and the resulted equation has only \((T - 1)\) observations on \( y^* \)s. It is possible to keep the first \( y_1 \) observation and use it directly for estimation as in

   \[
   (\sqrt{1 - \rho^2}) y_1 = (\sqrt{1 - \rho^2})(\beta_0 + x_1 \beta_1) + u_1^*
   \]

   where \( u_1^* = (\sqrt{1 - \rho^2}) \epsilon_1 \) has mean 0 and variance \( \sigma^2 \) (as the variance of \( u^* \)s).
• If $\rho$ were known, the OLS method can be applied to estimate $\beta$ using the transformed variables $x^*_t$ and $y^*_t$ for the regression equation

$$y^*_t = \beta_0 x^*_0 + \beta_1 x^*_1 + u^*_t \quad (**)$$

where $x^*_0 = \sqrt{1 - \rho^2}$ and $x^*_t = (1 - \rho)$ for $t = 2, \ldots, T$; and $x^*_1 = (\sqrt{1 - \rho^2})x_{11}$ and $x^*_t = (x_t - \rho x_{t-1})$ for $t = 2, \ldots, T$.

1) Because $u^*_t$ are no longer serially correlated, the classical assumptions are satisfied and the GLSE (i.e., OLS applied to the transformed equations) is the BLUE.

2) The GLSE has variance which can be smaller than that of the OLSE (when the latter is applied to the original data $(y,x)$).

3) As the dependent variable has changed after transformation, the GLS $R^2$ is not directly comparable to the OLS $R^2$.

• If $\rho$ is unknown, one has to estimate both $\beta$ and $\rho$.

1) There are many different methods (much complicated than the least squares method) which can estimate the regression model with serial correlated disturbances.

2) A feasible GLS method (two step method)

   (a) Estimate the regression equation $y = \beta_0 + \beta_1 x + e$ by the OLS method (ignoring the serial correlation problem).

   (b) Compute the OLS residual $e_t$ for all $t = 1, \ldots, T$. Estimate $\rho$ from the equation

   $$e_t = \rho e_{t-1} + u_t, \quad t = 2, \ldots, T$$

   by the method of OLS.

   (c) Use the estimate $\hat{\rho}$ of $\rho$ from the previous step to transform the data. Estimate either the transformed equation (*) with $T - 1$ observations, or estimate the equation (**) with $T$ observations.

3) An iterative method (Cochrane-Orcutt method)

   (a) Estimate $\rho$ by running a regression based on estimated residuals $\hat{e}_t$ of $e_t$:

   $$\hat{e}_t = \hat{\rho} \hat{e}_{t-1} + u_t, \quad t = 2, \ldots, T.$$ 

   (b) Use the estimate $\hat{\rho}$ from step (a) to do the quasi-difference. Estimate $\beta$ from the transformed equation (*).

   (c) With the estimate $\hat{\beta}$ from step (b) to construct the estimated residuals $\hat{e}_t$. Repeat the steps (a) and (b) until further iteration results in little change in the estimated $\hat{\rho}$ of $\rho$.

• An empirical example: estimate the model

$$Y_t = \beta_0 + \beta_1 PC_t + \beta_2 PB_t + \beta_3 YD_t + \epsilon_t$$

with (and without)

$$\epsilon_t = \rho \epsilon_{t-1} + u_t,$$

where
\(Y_t = \) per capita chicken consumption (in pounds) in year \(t\)

\(PC_t = \) the price of chicken (in cents per pound) in year \(t\)

\(PB_t = \) the price of beef (in cents per pound) in year \(t\)

\(YD_t = \) US per capita disposable income (in hundreds of dollars) in year \(t\).

1) OLS estimation (without taking into account serial correlation)

\[
\hat{Y}_t = 31.5 - 0.73PC_t + 0.11PB_t + 0.23YD_t
\]

\(t = -9.12 \quad (0.08) \quad 2.50 \quad 14.22 \quad (0.05) \quad (0.02) \quad (0.02)\)

\(\hat{R}^2 = 0.986, \quad n = 44 \) (annual 1951 – 1994), \( \text{DW} = 0.98\),

where \(\text{DW}\) stands for the Durbin-Watson statistic. From the table for DW statistic (with \(k = 3\) and \(n = 44\), \(d_L = 1.38\) and \(d_U = 1.67\). Because \(d = 0.98 < d_L\), we would reject the null hypothesis of no positive serial correlation.

2) Estimation with the Cochrane-Orcutt approach,

\[
\hat{Y}_t = 28.1 - 0.11PC_t + 0.09PB_t + 0.24YD_t
\]

\(t = -1.30 \quad (0.08) \quad 2.12 \quad 8.04 \quad (0.04) \quad (0.03)\)

\(\hat{R}^2 = 0.986, \quad n = 43 \) (annual 1951 – 1994), \( \hat{\rho} = 0.93\).

As \(\hat{\rho} = 0.93\), the estimation was actually run as \(Y_t^* = Y_t - 0.93Y_{t-1}\), \(PC_t^* = PC_t - 0.93PC_{t-1}\), etc.

Note \(n = 43\) because the first observation has used only as a lagged value for the quasi-difference.

3) The larger \(t\) values in the 1) OLS estimation could be over-estimated.