ECON 5360 Class Notes
Serial Correlation

1 Introduction

In this chapter, we focus on the problem of serial correlation (autocorrelation) within the multiple linear regression model. Throughout, we assume that all other classical assumptions are satisfied. Assume the model is

$$y_t = x_t'\beta + \epsilon_t \tag{1}$$

where

$$E(\epsilon\epsilon') = \sigma^2\Omega = \gamma_0 \begin{bmatrix} 1 & \rho_1 & \cdots & \rho_{n-2} & \rho_{n-1} \\ \rho_1 & 1 & \cdots & \rho_{n-3} & \rho_{n-2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \rho_{n-2} & \rho_{n-3} & \cdots & 1 & \rho_1 \\ \rho_{n-1} & \rho_{n-2} & \cdots & \rho_1 & 1 \end{bmatrix}, \tag{2}$$

$cov(\epsilon_t, \epsilon_{t-s}) = \gamma_s$ is called the autocovariance of the errors and $\rho_s = \gamma_s/\gamma_0$ is called the autocorrelation of the errors. Serial correlation is a common occurrence in time series data.

Consider an example of a macroeconomic consumption function

$$C_t = \beta_1 + \beta_2 t + \beta_3 Y_t + \epsilon_t$$

where $t = 1950, ..., 1985$, C_t is consumption and Y_t is income. A plot of the OLS residuals is attached.

2 Time Series Properties of a First-Order Autoregression

Assume that the error terms follow a first-order autoregressive (AR(1)) process

$$\epsilon_t = \rho\epsilon_{t-1} + \mu_t \tag{3}$$

where $\mu_t \sim iid(0, \sigma_\mu^2)$, $|\rho| < 1$ and $t = 1, ..., T$. Rewrite (3) using repeated substitutions

$$\begin{aligned} \epsilon_t &= \rho\epsilon_{t-1} + \mu_t \\ &= \rho(\rho\epsilon_{t-2} + \mu_{t-1}) + \mu_t = \rho^2\epsilon_{t-2} + \mu_t + \rho\mu_{t-1} \\ &= \vdots \\ &= \rho^T\epsilon_{t-T} + \sum_{j=0}^{T-1} \rho^j \mu_{t-j}. \end{aligned}$$

Letting $T \to \infty$ gives

$$\epsilon_t = \sum_{j=0}^{\infty} \rho^j \mu_{t-j}$$

which is called an infinite moving-average process (MA(∞)). We can now use the MA(∞) representation to calculate several moments of the distribution for ϵ_t.

- Mean.

$$E(\epsilon_t) = \sum_{j=0}^{\infty} \rho^j E(\mu_{t-j}) = 0.$$

- Variance.

$$\begin{aligned} \gamma_0 &= Var(\epsilon_t) = E(\epsilon_t^2) = E\left(\sum_{j=0}^{\infty} \rho^j \mu_{t-j}\right)^2 \\ &= \sigma_\mu^2 + \rho^2\sigma_\mu^2 + \rho^4\sigma_\mu^2 + \cdots \\ &= \sigma_\mu^2(1 + \rho^2 + \rho^4 + \cdots) \\ &= \sigma_\mu^2/(1-\rho^2). \end{aligned}$$

- Covariances.

$$\gamma_s = Cov(\epsilon_t, \epsilon_{t-s}) = E\left(\sum_{j=0}^{\infty} \rho^j \mu_{t-j}\right)\left(\sum_{j=0}^{\infty} \rho^j \mu_{t-s-j}\right) = \rho^s\sigma_\mu^2/(1-\rho^2) = \rho^s\gamma_0.$$

Since the mean does not depend on time and the covariance between error terms only depend on their distance between each other (and not t), we say that ϵ_t is a (weakly) covariance stationary process. This information can be substituted into (2) to give

$$\sigma^2\Omega(\rho) = \frac{\sigma_\mu^2}{(1-\rho^2)} \begin{bmatrix} 1 & \rho & \cdots & \rho^{T-2} & \rho^{T-1} \\ \rho & 1 & \cdots & \rho^{T-3} & \rho^{T-2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \rho^{T-2} & \rho^{T-3} & \cdots & 1 & \rho \\ \rho^{T-1} & \rho^{T-2} & \cdots & \rho & 1 \end{bmatrix}. \tag{4}$$

3 Ordinary Least Squares

We now examine several results related to OLS when autocorrelation is present in the model. It is useful to break these results into two parts – when the model has no lagged dependent variables and when it does.

3.1 Properties of OLS Estimators

3.1.1 No Lagged Dependent Variables

- $b = (X'X)^{-1}X'Y$ is unbiased and consistent.
- s^2 is a biased estimator of σ^2.
- s^2 is a consistent estimator of σ^2.
- $s^2(X'X)^{-1}$ is a biased and inconsistent estimate of $var(b) = \sigma^2(X'X)^{-1}X'\Omega X(X'X)^{-1}$.
- $b \overset{asy}{\sim} N(\beta, \frac{\sigma^2}{n}Q^{-1}\tilde{Q}Q^{-1})$ where $\text{plim}\frac{1}{n}(X'X) = Q$ and $\text{plim}\frac{1}{n}(X'\Omega X) = \tilde{Q}$.

3.1.2 Lagged Dependent Variables

Consider the following simple example

$$y_t = \beta y_{t-1} + \epsilon_t \tag{5}$$

where

$$\epsilon_t = \rho\epsilon_{t-1} + \mu_t \tag{6}$$

and μ_t is white noise. For this example, the OLS estimator is

$$b = \frac{\sum_{t=2}^{T} y_t y_{t-1}}{\sum_{t=2}^{T} y_{t-1}^2} = \beta + \frac{\sum_{t=2}^{T} \epsilon_t y_{t-1}}{\sum_{t=2}^{T} y_{t-1}^2}.$$

In the presence of a lagged dependent variable, the OLS estimates of β will be biased and inconsistent. The intuition is that y_{t-1} and ϵ_t are both directly influenced by ϵ_{t-1}. More specifically, y_{t-1} is influenced through equation (5) and ϵ_t is influenced through equation (6). The more formal argument using Slutsky's theorem and the fact that y_t and ϵ_t are covariance stationary gives

$$\begin{aligned}
\text{plim}(b) &= \beta + \text{plim}(\frac{1}{T}\sum_{t=2}^{T} y_{t-1}^2)^{-1}\text{plim}(\frac{1}{T}\sum_{t=2}^{T} \epsilon_t y_{t-1}) \\
&= \beta + (\frac{\sigma_\mu^2}{(1-\rho^2)(1-\beta^2)})^{-1}(\frac{\rho\sigma_\mu^2}{(1-\rho^2)(1-\rho\beta)}) \\
&= \beta + \frac{\rho(1-\beta^2)}{(1-\rho\beta)}.
\end{aligned}$$

To summarize, we know the following about OLS in the presence of lagged dependent variables:

- $b = (X'X)^{-1}X'Y$ is biased and inconsistent.
- s^2 is a biased and inconsistent estimator of σ^2.
- $s^2(X'X)^{-1}$ is a biased and inconsistent estimate of $var(b) = \sigma^2(X'X)^{-1}X'\Omega X(X'X)^{-1}$.

3.2 Autocorrelation-Corrected Standard Errors with OLS

For autocorrelation, Newey and West have developed a consistent estimator of $var(b) = \sigma^2(X'X)^{-1}X'\Omega X(X'X)^{-1}$ in the same spirit as White's standard error correction with heteroscedasticity. The Newey-West estimator is also robust to different forms of autocorrelation (i.e., you do not need to directly specify the autoregressive process for ϵ_t). Under the Newey-West estimator, we calculate

$$S_* = \frac{1}{T}\sum_{t=1}^{T} e_t^2 x_t x_t' + \frac{1}{T}\sum_{j=1}^{L}\sum_{t=j+1}^{T}(1 - \frac{j}{L+1})e_t e_{t-j}[x_t x_{t-j}' + x_{t-j}x_t']$$

where L is chosen large enough that the residual correlations are insignificant. Newey and West show that S_* will converge in probability to $\frac{\sigma^2}{n}X'\Omega X$, where the e_t are the OLS residuals.

4 Testing for Autocorrelation

As in the case of heteroscedasticity, the tests below are based on the OLS residuals. This makes sense, at least asymptotically, because $b \xrightarrow{p} \beta$.

4.1 Graphical Test

As a first step, it is often useful to graph e_t against t and see if the residuals appear to be random noise. See the attached graph of consumption residuals and Gauss example #5.

4.2 Durbin-Watson Test

This is the most widely used autocorrelation test. It is used to test for first-order autocorrelation (i.e., $\epsilon_t = \rho\epsilon_{t-1} + \mu_t$). The null hypothesis is H_0: $\rho = 0$ and the alternative hypothesis is H_A: $\rho \neq 0$. Let's begin by rearranging the test statistic:

$$\begin{aligned} d &= \frac{\sum_{t=2}^{T}(e_t - e_{t-1})^2}{\sum_{t=1}^{T} e_t^2} = \frac{\sum_{t=2}^{T}(e_t^2 - 2e_t e_{t-1} + e_{t-1}^2)}{\sum_{t=1}^{T} e_t^2} \\ &= \frac{2\sum_{t=1}^{T} e_t^2 - 2\sum_{t=2}^{T} e_t e_{t-1} - e_1^2 - e_T^2}{\sum_{t=1}^{T} e_t^2} \\ &= 2(1 - r_1) - \Delta \end{aligned}$$

where

$$r_j = \frac{\sum_{t=j+1}^{T} e_t e_{t-j}}{\sum_{t=1}^{T} e_t^2}$$

is an estimate of the j^{th}-order autocorrelation coefficient and

$$\Delta = \frac{e_1^2 + e_T^2}{\sum_{t=1}^{T} e_t^2}$$

is a term that goes to zero as $T \to \infty$. Therefore, in the limit, we know $d = 2(1 - r_1)$. If H_0 is true, we would expect $r_1 = 0$ and $d = 2$. If $\rho = 1$, then we would expect $d = 0$. If $\rho = -1$, then we would expect $d = 4$. A couple of notes.

- The exact distribution for the d statistic depends upon X, and as a result, a unique set of critical values does not exist. Durbin and Watson have, however, developed lower and upper bounds for the true but unknown critical values. If the d statistic falls in between the lower and upper bounds for the critical value, no conclusion can be reached.

- Positive autocorrelation (i.e., $\rho > 0$) is much more common than negative autocorrelation, so often the test is one-tailed with the critical values at the lower end of the distribution.

4.2.1 Durbin's h test

Not surprisingly, the DW test does not work in the presence of lagged-dependent variables because the OLS estimates of β are biased and inconsistent. Durbin has developed an alternative. The test statistic is

$$h = r_1 \sqrt{T/(1 - T s_c^2)}$$

where s_c^2 is the estimated variance of the coefficient on y_{t-1}. The statistic h has an asymptotic standard normal distribution and can be used to test the same H_0 as in the DW test.

4.3 Lagrange Multiplier Test

The disadvantage of the DW test is that it has an inconclusive region and only works for AR(1) processes. The LM test helps resolve these issues. The hypotheses are

$$\begin{aligned} H_0 &: \quad \text{no autocorrelation} \\ H_A &: \quad AR(p) \text{ or } MA(p). \end{aligned}$$

The first step is to run the following regression

$$e_t = x_t \gamma' + \rho_1 e_{t-1} + \rho_2 e_{t-2} + \cdots + \rho_p e_{t-p} + \nu_t.$$

The test statistic is $LM = TR^2$ which has an asymptotic chi-square distribution with p degrees of freedom. There is a tradeoff associated with the choice of p. Choosing too large of a p can cause the test to lose power (i.e., lead to Type II errors). Choosing too small of a p may miss higher-order autocorrelation.

4.4 Box-Pierce Q Test

The Box-Pierce Q test is similar to the LM test but it does not control for X. The test statistic is

$$Q = T \sum\nolimits_{j=1}^{p} r_j^2$$

which is asymptotically chi-square with p degrees of freedom. A slight variation of the Box-Pierce Q test was suggested by Ljung and Box

$$Q' = T(T+2) \sum\nolimits_{j=1}^{p} \frac{r_j^2}{T-j}.$$

4.5 Gauss Example (cont.)

We now perform the three tests for autocorrelation using the U.S. consumption function example. See Gauss example #6 for the results.

5 Generalized Least Squares

5.1 Ω is Known

The efficient estimator for the model in equations (1) and (2) is

$$\begin{aligned}\hat{\beta} &= (X'\Omega^{-1}X)^{-1}(X'\Omega^{-1}Y) \\ &= (X'P'PX)^{-1}(X'P'PY).\end{aligned}$$

We need to calculate the transformation matrix P such that the transformed errors are white noise. Toward that end, define the lag operator L to be $L^j X_t = X_{t-j}$. The appropriate transformation is

$$(1-\rho L)Y = (1-\rho L)X\beta + (1-\rho L)\epsilon \Rightarrow Y_* = X_*\beta + \mu$$

where

$$Y_* = \begin{bmatrix} y_1 - \rho y_0 \\ y_2 - \rho y_1 \\ \vdots \\ y_T - \rho y_{T-1} \end{bmatrix} \quad X_* = \begin{bmatrix} x_1 - \rho x_0 \\ x_2 - \rho x_1 \\ \vdots \\ x_T - \rho x_{T-1} \end{bmatrix} \quad \epsilon_* = \mu = \begin{bmatrix} \epsilon_1 - \rho \epsilon_0 \\ \epsilon_2 - \rho \epsilon_1 \\ \vdots \\ \epsilon_T - \rho \epsilon_{T-1} \end{bmatrix}.$$

The problem is how to transform the first observation since y_0 and x_0 are not observed. One solution is to treat y_1 and x_1 as starting values and multiply the entire first row by $\sqrt{1-\rho^2}$ so that $\epsilon_1^* \sim N(0, \sigma_\mu^2)$. This implies that the transformation matrix will be

$$P = \begin{bmatrix} \sqrt{1-\rho^2} & 0 & \cdots & 0 & 0 \\ -\rho & 1 & & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & & 1 & 0 \\ 0 & 0 & \cdots & -\rho & 1 \end{bmatrix}. \tag{7}$$

Technically, this means that the transformed model has no constant since the transformed constant has a different first value. Also, remember that the transformation matrix P above is only valid for AR(1) autocorrelation processes. Higher-order process involve more complicated P.

5.1.1 Maximum Likelihood Estimation

Start by writing the transformed model as

$$y_t = \rho y_{t-1} + x_t'\beta - \rho x_{t-1}'\beta + \mu_t.$$

The likelihood (joint probability) function can then be written as

$$L(\theta) = f(y_1) f(y_2|y_1) \cdots f(y_T|y_{T-1})$$

where we iteratively use the definition of a conditional distribution

$$f(x_2|x_1) = \frac{f(x_1, x_2)}{f(x_1)}.$$

The log likelihood function is then

$$\ln L(\theta) = \ln f(y_1) + \sum\nolimits_{t=2}^{T} \ln f(y_t|y_{t-1}).$$

Assuming that f is the normal pdf, we get

$$\ln L(\theta) = \left(-\frac{T}{2}[\ln(2\pi) + \ln(\sigma_\mu^2)] - \frac{1}{2\sigma_\mu^2}\sum\nolimits_{t=1}^{T}\mu_t^2\right) + \frac{1}{2}\ln(1-\rho^2) \tag{8}$$

where $\theta = \{\beta, \sigma_\mu^2, \rho\}$ and the last term in (8) is included to account for the first observation. Assuming that $\Omega(\rho)$ is known, the ML estimator is the familiar GLS estimator

$$\begin{aligned} \hat{\beta} &= (X'\Omega^{-1}X)^{-1}(X'\Omega^{-1}Y) \text{ and} \\ \hat{\sigma}_\mu^2 &= (e'\Omega^{-1}e)/T. \end{aligned}$$

5.2 Ω is Unknown

Next we consider two types of estimators when Ω is unknown – two step and maximum likelihood estimators.

5.2.1 Two-Step Estimators

Both estimators are asymptotically efficient with one iteration. Further iterations are optional and may help small-sample performance.

1. Prais and Winsten. This is a type of feasible GLS. The estimator works as follows
   - Step #1. Use r_1 as an estimate of ρ. Transform the data according to (7).
   - Step #2. Run OLS on transformed data (i.e., $\hat{\beta} = (X_*'X_*)^{-1}X_*'Y_*$).
2. Cochrane-Orcutt. Same as above but ignore the first observation.

5.2.2 Maximum Likelihood Estimators

1. Brute force. Maximize (8) using one of many nonlinear optimization algorithms, such as Newton-Raphson.
2. Hildreth and Lu. Use a grid search to find an estimate of ρ that maximize (8) for the implied estimates of β and σ_μ^2.

5.3 Gauss Application (cont.)

We continue with the consumption function example and contrast three different estimators – Prais-Winsten, Cochrane-Orcutt and maximum likelihood. See Gauss example #6 for further details.

GAUSS Fri Nov 14 13:12:43 2003