Economics 522A, Spring 2007

Lecture Note 17: Serial Correlation Part I

1 Introduction

Consider the generalized classical regression model:

$$E[y|X] = X\beta, \quad V[y|X] = \Omega.$$

If the off-diagonal elements of Ω are nonzero, this means that for some $i \neq j$,

$$Cov[y_i, y_j|X] \neq 0.$$

We say there is serial correlation among the observations.

Serial correlation is often encountered when working with time-ordered observations, for example observations on macroeconomic variables. When the data are a time series, we often use subscripts $t = 1, \ldots, T$ to denote the observations, rather than $i = 1, \ldots, n$. There is nothing special about using t in place of i, except as a mnemonic to help us remember that the data have an interpretation as time series.

Example: Phillips Curve

Suppose

$$y_t = \text{ change in inflation between period } t \text{ and } t-1.$$

$$U_t = \text{ unemployment rate at time } t-1,$$

$$z_t = \text{ other macroeconomic variables related to supply shocks.}$$

We might suppose that

$$E[y_t|X] = \beta_1 + \beta_2 U_t + \beta_3' z_t,$$

where X denotes $U_1, z_1, \ldots, U_T, z_T$. It would be very plausible that inflation changes are serially correlated, even after controlling for unemployment and the variables in z_t. One way we could assess this possibility, is to run the OLS regression of y_t on a constant, U_t, and z_t. Then form the OLS residuals $e_t = y_t - \hat{\beta}_1 - \hat{\beta}_2 U_t - \hat{\beta}_3' z_t$.

We could then calculate the sample correlation between e_{t+1} and e_t, for example, and see if it is very different from 0.

2 Serial Correlation: AR(1) model

We will start with a specific model for the correlation across observations. Suppose that for observations $t = 1, 2, \ldots, T$,

$$y_t = x_t'\beta + \epsilon_t,$$

and that, for $t = 2, \ldots, T$,

$$\epsilon_t = \phi\epsilon_{t-1} + v_t,$$

where the v_t are IID $N(0, \sigma_v^2)$, independent of all other variables. This says that the ϵ_t is a Gaussian AR(1) stochastic process. We will assume that $|\phi| < 1$. (Recall the discussion of stochastic processes in Econ 520.)

We also need to specify the distribution of ϵ_1, the initial observation's disturbance term. Let us assume that

$$\epsilon_1|X \sim N\left(0, \frac{\sigma_v^2}{1-\phi^2}\right).$$

We'll explain why we chose that particular expression for the variance of ϵ_1 in a few moments. (No pun intended.)

To understand the structure of the disturbance terms, we can write:

$$\begin{aligned}
\epsilon_2 &= \phi\epsilon_1 + v_2 \\
\epsilon_3 &= \phi\epsilon_2 + v_3 \\
&= \phi(\phi\epsilon_1 + v_2) + v_3 \\
&= \phi^2\epsilon_1 + \phi v_2 + v_3
\end{aligned}$$

Similarly, by substituting recursively,

$$\begin{aligned}
\epsilon_4 &= \phi^3\epsilon_1 + \phi^2 v_2 + \phi v_3 + v_4 \\
&\vdots \\
\epsilon_t &= \phi^{t-1}\epsilon_1 + \phi^{t-2}v_2 + \phi^{t-3}v_3 + \cdots + v_t.
\end{aligned}$$

Note also that since $|\phi| < 1$ by assumption, $\phi^t \to 0$ as $t \to \infty$. So eventually, the influence of the initial term ϵ_1 on ϵ_t becomes very small.

Now, let us reconsider ϵ_1. Suppose that AR(1) process has been running for a very long time in the past, that is, for observations $t = 0, -1, -2, -3, \ldots$. Then we could write

$$\epsilon_1 \approx v_1 + \phi v_0 + \phi^2 v_{-1} + \cdots$$

Since it is a linear combination of IID $N(0, \sigma_v^2)$ terms, it will be normal, with mean

$$E[\epsilon_1] = E[v_1 + \phi v_0 + \phi^2 v_{-1} + \cdots] = 0$$

and variance

$$\begin{aligned}
V[\epsilon_1] &= V[v_1 + \phi v_0 + \phi^2 v_{-1} + \cdots] \\
&= \sigma_v^2 + \phi^2\sigma_v^2 + \phi^4\sigma_v^2 + \cdots \\
&= \frac{\sigma_v^2}{1-\phi^2}.
\end{aligned}$$

So if we want to assume that ϵ_1 is drawn from the "long-run" marginal distribution of the ϵ_t process, then our choice for its distribution is the appropriate one.

Now, let us work out what this model implies about the covariance matrix Ω. We have already specified the mean and variance of ϵ_1. Next, consider ϵ_2:

$$E[\epsilon_2] = E[\phi\epsilon_1 + v_2] = 0.$$

$$\begin{aligned} V[\epsilon_2] &= V[\phi\epsilon_1 + v_2] \\ &= \phi^2 V[\epsilon_1] + V[v_2] \\ &= \phi^2 \frac{\sigma_v^2}{1-\phi^2} + \sigma_v^2 \\ &= \frac{\sigma_v^2}{1-\phi^2}. \end{aligned}$$

By the same argument, for all $t = 1, \ldots, T$,

$$E[\epsilon_t] = 0, \quad V[\epsilon_t] = \frac{\sigma_v^2}{1-\phi^2}.$$

We also need to figure out the covariance terms in Ω:

$$\begin{aligned} Cov(\epsilon_t, \epsilon_{t+1}) &= Cov(\epsilon_t, \phi\epsilon_t + v_{t+1}) \\ &= E[(\epsilon_t)(\phi\epsilon_t + v_{t+1})] \\ &= E[\phi\epsilon_t^2 + \epsilon_t v_{t+1}] \\ &= \phi E[\epsilon_t^2] + 0 \\ &= \phi V[\epsilon_t]. \end{aligned}$$

Similarly,

$$\begin{aligned} Cov(\epsilon_t, \epsilon_{t+2}) &= E[\epsilon_t(\phi^2\epsilon_t + \phi v_{t+1} + v_{t+2})] \\ &= E[\phi^2\epsilon_t^2] \\ &= \phi^2 V[\epsilon_t] \end{aligned}$$

So, letting $\sigma^2 = \sigma_v^2/(1-\phi^2)$, we can write

$$V[y|X] = \Omega = \sigma^2 \begin{pmatrix} 1 & \phi & \phi^2 & \cdots & \phi^{T-1} \\ \phi & 1 & \phi & \cdots & \phi^{T-2} \\ \phi^2 & \phi & 1 & & \\ \vdots & \vdots & & \ddots & \\ \phi^{T-1} & \phi^{T-2} & \cdots & & 1 \end{pmatrix}.$$

(Note: by assumption, the ϵ_t process is completely independent of X, so the conditional variances and covariances are equal to the unconditional ones we calculated above.)

So the covariance between observations j periods apart should decline geometrically as j increases, if the AR(1) model generates the data. We could use the OLS residuals and calculate sample covariances, and see if they seem to fit this predicted pattern.

The AR(1) specification is not the only possible form of serial correlation, of course. The disturbance terms could have an AR(2) structure, an MA(1) structure, or many other possible forms of serial correlation, which would imply different forms for the matrix Ω.

3 Properties of OLS under Serial Correlation

We have already shown that OLS is unbiased and worked out its variance matrix under general forms of Ω. So those results still apply here. Also, we showed for general Ω that GLS is the minimum variance linear unbiased estimator for β.

It can be shown that OLS is consistent and asymptotically normal. However, there are some slight complications here. For example, consider our usual proof of consistency. First, we would write

$$\begin{aligned}
\hat{\beta} &= \left(\frac{1}{T}\sum_{t=1}^{T} x_t x_t'\right)^{-1} \frac{1}{T}\sum_{t=1}^{T} x_t y_t \\
&= \left(\frac{1}{T}\sum_{t=1}^{T} x_t x_t'\right)^{-1} \frac{1}{T}\sum_{t=1}^{T} x_t (x_t'\beta + \epsilon_t) \\
&= \beta + \left(\frac{1}{T}\sum_{t=1}^{T} x_t x_t'\right)^{-1} \frac{1}{T}\sum_{t=1}^{T} x_t \epsilon_t
\end{aligned}$$

At this point, we'd want to use the law of large numbers to show that the last term converges to 0. However, since we want to think of the process (x_t, y_t) as being correlated over time, we cannot assume that they are IID. So our usual LLN and CLT results cannot be used here.

Fortunately, there are extensions of the LLN and CLT that apply when the observations are serially dependent. One thing you have to be careful about is that

$$V\left[\frac{1}{T}\sum_{t=1}^{T} x_t \epsilon_t\right] \neq \frac{1}{T^2}\sum_{t=1}^{T} V[x_t \epsilon_t],$$

because there are additional covariance terms.

We need the left hand side of the previous expression to go to 0 as $T \to \infty$. This requires that the covariances across observations "die off" sufficiently quickly as observations become far apart in time. See Ruud for more detail on this point, and on extensions of the classic LLN and CLT to dependent data.