Economics 522A, Spring 2007

Lecture Note 18: Serial Correlation Part II (corrected 3/29/07)

1 Estimating the Variance of OLS

We want to understand the variance of OLS under serial correlation and develop some ways to consistently estimate the variance, so that we can form standard errors, confidence intervals, and valid hypothesis testing procedures.

It's useful to start with a general situation. Suppose that $V[y|X] = \Omega$, and write

$$\Omega = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1T} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2T} \\ \vdots & & \ddots & \\ \sigma_{T1} & \sigma_{T2} & \cdots & \sigma_{TT} \end{pmatrix},$$

where $\sigma_{st} = E[\varepsilon_s \varepsilon_t | X]$. Of course, any variance matrix must be symmetric, so we must have $\sigma_{st} = \sigma_{ts}$.

Recall that

$$V[\hat{\beta}|X] = (X'X)^{-1} X' \Omega X (X'X)^{-1}.$$

So a key piece of the variance of OLS is $X'\Omega X$. Let's work this out explicitly. First, consider ΩX. We can write

$$\begin{aligned} \Omega X &= \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1T} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2T} \\ \vdots & & \ddots & \\ \sigma_{T1} & \sigma_{T2} & \cdots & \sigma_{TT} \end{pmatrix} \begin{pmatrix} x_1' \\ x_2' \\ \vdots \\ x_T' \end{pmatrix} \\ &= \begin{pmatrix} \sigma_{11} x_1' + \sigma_{12} x_2' + \cdots + \sigma_{1T} x_T' \\ \sigma_{21} x_1' + \sigma_{22} x_2' + \cdots + \sigma_{2T} x_T' \\ \vdots \\ \sigma_{T1} x_1' + \sigma_{T2} x_2' + \cdots + \sigma_{TT} x_T' \end{pmatrix}. \end{aligned}$$

Therefore,

$$\begin{aligned} X'\Omega X &= [x_1, \ldots, x_T]\Omega X \\ &= x_1 \sigma_{11} x_1' + x_1 \sigma_{12} x_2' + \cdots + x_1 \sigma_{1T} x_T' \\ &\quad + x_2 \sigma_{21} x_1' + \cdots \\ &\quad \vdots \\ &\quad + x_T \sigma_{T1} x_1' + \cdots + x_T \sigma_{TT} x_T'. \end{aligned}$$

Collecting terms, we can write

$$X'\Omega X = \sum_{t=1}^{T} x_t \sigma_{tt} x_t' + \sum_{t=2}^{T} x_t \sigma_{t,t-1} x_{t-1}' + \sum_{t=2}^{T} x_{t-1} \sigma_{t-1,t} x_t'$$

$$
\begin{aligned}
&+\sum_{t=3}^{T} x_t \sigma_{t,t-2} x'_{t-2} + \sum_{t=3}^{T} x_{t-2} \sigma_{t-2,t} x'_t \\
&+\cdots
\end{aligned}
$$

In order to estimate this, we need to restrict it somewhat. Suppose for example that the all the covariances of order greater than $p < T$ are equal to 0. So, for example, $\sigma_{1,p+1} = 0$. Then the expression above will be truncated.

For $j = 0, 1 \ldots, p$, let

$$\hat{\Lambda}_{Tj} = \frac{1}{T} \sum_{t=j+1}^{T} x_t e_t e_{t-j} x'_{t-j},$$

where the e_t are the OLS residuals. Let

$$\hat{\Lambda}_T = \hat{\Lambda}_{T0} + \sum_{j=1}^{p} \left(\hat{\Lambda}_{Tj} + \hat{\Lambda}'_{Tj} \right).$$

Under a law of large numbers argument, $\hat{\Lambda}_T$ should approximate

$$\frac{1}{T} X' \Omega X.$$

This is a variant of the Eicker-White estimator we studied in the section on heteroskedasticity.

In order for this to work, we generally need p to be fairly small relative to the sample size T. What if this is not the case, as in our AR(1) example? Hansen (1982) suggested using a finite p, but letting it grow as a function of sample size T. Under some conditions, most importantly that the covariances decline as the order increases, this can be shown to lead to a consistent estimator.

Hansen's estimator can sometimes fail to be positive definite, if p is too large. A popular alternative is the Newey-West (1987) estimator, which has a similar form, but downweights the higher-order terms:

$$\hat{\Lambda}_T = \hat{\Lambda}_{T0} + \sum_{j=1}^{p} \left(1 - \frac{j}{p+1} \right) \left(\hat{\Lambda}_{Tj} + \hat{\Lambda}'_{Tj} \right).$$

2 Estimating Autoregressions

Consider a slightly different estimation problem. Suppose we assume that for $t = 2, \ldots, T$,

$$y_t = \beta_1 + \beta_2 y_{t-1} + \epsilon_t,$$

where

$$\epsilon_t | y_1, \ldots, y_{t-1} \overset{\text{ind.}}{\sim} N(0, \sigma^2).$$

This is a pure autoregressive model for the y_t.

We could write this in stacked form as:

$$\begin{pmatrix} y_2 \\ y_3 \\ \vdots \\ y_T \end{pmatrix} = \begin{pmatrix} 1 & y_1 \\ 1 & y_2 \\ \vdots & \vdots \\ 1 & y_{T-1} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} \epsilon_2 \\ \epsilon_3 \\ \vdots \\ \epsilon_T \end{pmatrix}$$

or

$$\begin{pmatrix} y_2 \\ \vdots \\ y_T \end{pmatrix} = X\beta + \epsilon.$$

Since this has the usual form for the classical regression model, we might be tempted to say that OLS is unbiased. However, in fact, OLS is <u>biased</u> here.

To understand the problem, recall that our proof of unbiasedness of OLS relied upon $E[y|X] = X\beta$. However, here,

$$E[y|X] \neq X\beta.$$

The problem is that X contains lagged values of y, in particular $y_1, y_2, \ldots, y_{T-1}$. So for example,

$$E[y_3|X] = y_3,$$

because X contains y_3.

It is still possible to show that OLS is consistent, and if the process for y_t is stationary, that OLS is asymptotically normal.

3 Clustering

Sometimes, data are "grouped" or "clustered." For example, we might have a sample of individuals, where individuals reside in different cities. We could think of the individuals as being grouped by city.

Data with a natural grouping structure are called *panel data*. A special case of panel data is *longitudinal data*, where units are followed over time. For example, a data set called the Panel Study of Income Dynamics follows families over time. The family's income and other socioeconomic variable are collected in every year.

In another data set, we might have a sample of firms, and observe multiple individual workers in each firm. Again, there is a grouping structure, by firm.

We will use $i = 1, \ldots, n$ to denote the groups, and $t = 1, \ldots, T$ to denote individual observations within the group. (For simplicity, we will assume that each group i has the same number T of observations.) So, for example, y_{it} will be the observation on y for group i, unit t. The notation is natural in the case of longitudinal data, where i might be individual, and y_{it} might be income for individual i in year t. We will talk about individuals and years for the rest of the note just to fix ideas, but keep in mind that the meaning of i and t could be different depending on the specific application at hand.

In panel data, we might have regressor variables at various levels. Let w_i denote regressors that vary across individuals but not over time, and let z_{it} denote regressors that vary across individuals and time. Suppose that

$$E[y_{it}|w_1, \ldots, w_n, z_{11}, \ldots, z_{nT}] = w_i'\lambda + z_{it}'\gamma.$$

We can expression this in vector-matrix form as follows. Stack the observations as:

$$y = \begin{pmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1T} \\ y_{21} \\ \vdots \\ y_{2T} \\ \vdots \\ y_{n1} \\ \vdots \\ y_{nT} \end{pmatrix}, \quad X = \begin{pmatrix} w_1' & z_{11}' \\ w_1' & z_{12}' \\ \vdots & \vdots \\ w_1' & z_{1T}' \\ w_2' & z_{21}' \\ \vdots & \vdots \\ w_2' & z_{2T}' \\ \vdots & \vdots \\ w_n' & w_{n1}' \\ \vdots & \vdots \\ w_n' & w_{nT}' \end{pmatrix}, \quad \beta = \begin{pmatrix} \lambda \\ \gamma \end{pmatrix}.$$

Then we can write

$$E[y|X] = X\beta.$$

So this fits into our generalized linear regression model form. If we assume that

$$V[y|X] = \sigma^2 I_{nT},$$

then we would have the classical regression model. However, we might think that there would be correlation across the different observations for a given individual.

Suppose we assume that

$$\begin{aligned} V[y_{it}|X] &= \sigma^2, \quad \forall i = 1, \ldots, n, \\ Cov[y_{it}, y_{is}|X] &= \rho\sigma^2, \quad \forall i, \forall t \neq s, \\ Cov[y_{it}, y_{js}|X] &= 0, \quad \forall i \neq j, \forall t, s. \end{aligned}$$

So the covariance matrix of y given X will have a block-diagonal form

$$\Omega = \begin{pmatrix} \Sigma & 0 & \cdots & 0 \\ 0 & \Sigma & \cdots & 0 \\ \vdots & & \ddots & 0 \\ 0 & & & \Sigma \end{pmatrix},$$

where Σ is $T \times T$ with

$$\Sigma = \sigma^2 \begin{pmatrix} 1 & \rho & \rho & \cdots & \rho \\ \rho & 1 & \rho & \cdots & \rho \\ \rho & \rho & 1 & & \\ \vdots & \vdots & & \ddots & \\ \rho & \rho & \cdots & & 1 \end{pmatrix}.$$

One way to generate this "equicorrelated" variance matrix is to assume that the disturbance term has an "error-components" structure. Suppose we can write

$$y_{it} = w_i'\lambda + z_{it}'\gamma + v_i + u_{it},$$

where

$$\begin{aligned} E[v_i|X] &= 0 \\ E[u_{it}|X] &= 0 \\ V[v_i|X] &= \sigma_v^2 \\ V[u_{it}|X] &= \sigma_u^2, \end{aligned}$$

and all the v_i and u_{it} are uncorrelated conditional on X. Then let $\varepsilon_{it} = v_i + u_{it}$, and notice that

$$V[\varepsilon_{it}|X] = \sigma_v^2 + \sigma_u^2,$$

and that for $t \neq s$,

$$\begin{aligned} Cov(\varepsilon_{it}, \varepsilon_{is}|X) &= E[\varepsilon_{it}\varepsilon_{is}|X] \\ &= E[(v_i + u_{it})(v_i + u_{is})|X] \\ &= \sigma_v^2. \end{aligned}$$

Also, for $i \neq j$, $Cov(\varepsilon_{it}, \varepsilon_{js}|X) = 0$. So the variance matrix of y given X will have the equicorrelated form with $\sigma^2 = \sigma_v^2 + \sigma_u^2$ and $\rho = \sigma_v^2/(\sigma_v^2 + \sigma_u^2)$.

Since the variance matrix is parametrized by only two parameters, σ^2 and ρ, we could estimate these parameters using the residuals from a preliminary OLS regression. We could then form standard errors using the estimate of Ω, or we could do FGLS.

We could allow for a more general covariance structure. For example, we could assume that

$$\Omega = \begin{pmatrix} \Sigma_1 & & & 0 \\ & \Sigma_2 & & \\ & & \ddots & \\ 0 & & & \Sigma_n \end{pmatrix}.$$

If n is large, relative to T, then this variance matrix could be estimated from the OLS residuals, using a strategy similar to the Eicker-White approach for heteroskedastic data (and the approach we outlined for time-series data above).

4 Correlated Random Effects and Fixed Effects

In the error-components model of the previous section, v_i is sometimes called a "random effect." Very important: be careful about interpreting the error-components model. The v_i is an individual-specific term, but we are assuming that $E[v_i|X] = 0$, which implies that v_i is *uncorrelated* with the regressors. In some situations, we have a specific interpretation of v_i, and this interpretation might require that v_i is allowed to be correlated with the regressors. If this is the case, then

$$E[y_{it}|X] = w_i'\lambda + z_{it}'\gamma + E[v_i|X],$$

which is a different model. OLS of y_{it} on w_i and z_{it} will not necessarily be unbiased for λ and γ, because the conditional mean of y_{it} is different from what we originally assumed.

If this is the case, we need to use different methods to estimate β, generally called "correlated random effects" and "fixed effects" methods. These methods will be discussed in more detail in 522B.