WARNING!!!

THIS DOCUMENT CONTAINS TYPOS. THESE TYPOS WERE CORRECTED DURING THE LECTURE, BUT THEY MAY NOT HAVE BEEN CORRECTED IN THIS DOCUMENT

4.5 The Consequences of Serial Correlation

What are the consequences if we use least squares to estimate a model in which the error terms are in fact serially correlated? Let us suppose we estimate the model

$$Y = X\beta + u, \quad E(uu') = \sigma^2 I, \tag{117}$$

when the data-generating process is actually

$$Y = X\beta_0 + u, \tag{118}$$

$$u_t = \rho u_{t-1} + \epsilon_t, \quad \epsilon \sim IID(0, \omega_0^2).$$

The OLS estimator is

$$\hat{\beta} = (X'X)^{-1}X'Y, \tag{119}$$

which under the GDP is equal to

$$\hat{\beta} = (X'X)^{-1}X'(X\beta_0 + u) \tag{120}$$

$$= \beta_0 + (X'X)^{-1}X'u. \tag{121}$$

Provided that X is exogenous, $\hat{\beta}$ will still be unbiased, because the fact that the u_t's are serially correlated does not prevent $E(X'u)$ from being zero. If X is not exogenous, $\hat{\beta}$ will be consistent as long as $plim(n^{-1}X'u)$ is equal to zero.

Inferences about β will not be correct, however. Assuming that X is exogenous, we see that

$$E(\hat{\beta} - \beta_0)(\hat{\beta} - \beta_0)' = E\left[(X'X)^{-1}X'uu'X(X'X)^{-1}\right] \tag{122}$$

$$= (X'X)^{-1}X'\Omega_0 X(X'X)^{-1}, \tag{123}$$

where Ω_0 is the matrix defined by the GDP.

Except in special cases, it is not possible to say whether the incorrect standard error estimates obtained using OLS will be larger or smaller than the correct ones obtained by taking the square roots of the diagonal elements in the usual fashion. However, analysis of special cases suggests that for values of ρ greater than 0 (the most commonly encountered cases) the incorrect OLS standard errors are usually too small—see Nicholls and Pagan (1977), Sathe and Vinod (1974), and Vinod (1976).

4.6 Moving Average and ARMA Processes

Autoregressive processes are not the only way to model stationary time series. The other basic type of stochastic process is the **moving average**, or **MA**, process. The simplest moving average process is the **first-order moving average**, or **MA(1)**, process

$$u_t = \epsilon_t + \alpha_1 \epsilon_{t-1}, \quad \epsilon_t \sim IID(0, \omega^2), \tag{124}$$

in which the error u_t is literally a moving average of two successive innovations, ϵ_t and ϵ_{t-1}. Thus ϵ_t affects both u_t and u_{t+1} but does not affect u_{t+j} for $j > 1$.

Year-on-year inflation is an example of a MA process.

The more general MA(q) process may be written as

$$u_t = \epsilon_t + \alpha_1 \epsilon_{t-1} + \alpha_2 \epsilon_{t-2} + \cdots + \alpha_q \epsilon_{t-q}, \quad \epsilon_t \sim IID(0, \omega^2). \tag{125}$$

Finite-order MA processes are necessarily stationary, since each u_t is a weighted sum of a finite number of innovations $\epsilon_t, \epsilon_{t-1} \cdots$. Thus we do not have to impose stationarity conditions.

We do, however, have to impose an **invertibility condition** if we want α to be identifiable from data. In the MA(1) case, this condition is that $|\alpha_1| \leq 1$.

The reason we need an invertibility condition is that otherwise there will, in general, be more than one value of α that will yield any observed behavior pattern of the u_t's.

For example, the MA(1) process with $\alpha_1 = \gamma$, $-1 < \gamma < 1$, can be shown to be indistinguishable form an MA(1) process with $\alpha_1 = \dfrac{1}{\gamma}$.

It is straightforward to calculate the covariance matrix for a moving average process. for example, in the MA(1) case the variance of u_t is evidently

$$\sigma^2 = E(e_t + \alpha_1 \epsilon_{t-1})^2 \tag{126}$$

$$= \omega^2 + \alpha_1^2 \omega^2 \tag{127}$$

$$= (1 + \alpha_1^2)\omega^2, \tag{128}$$

the covariance of u_t and u_{t-1} is

$$E(\epsilon_t + \alpha_1 \epsilon_{t-1})(\epsilon_{t-1} + \alpha_1 \epsilon_{t-2}) = \alpha_1 \omega^2, \tag{129}$$

and the covariance of u_t and u_{t-j} for $j > 1$ is zero. Thus the covariance matrix of U is

$$\omega^2 \begin{bmatrix} 1+\alpha_1^2 & \alpha_1 & 0 & \cdots & 0 & 0 & 0 \\ \alpha_1 & 1+\alpha_1^2 & \alpha_1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \alpha_1 & 1+\alpha_1^2 & \alpha_1 \\ 0 & 0 & 0 & \cdots & 0 & \alpha_1 & 1+\alpha_1^2 \end{bmatrix} \tag{130}$$

Notice that the correlation between successive error terms varies only between -0.5 and 0.5, since those are the smallest and largest possible values of $\dfrac{\alpha_1}{1+\alpha_1^2}$, achieved when $\alpha_1 = -1$ and $\alpha_1 = 1$, respectively. It is thus evident that an MA(1) process cannot be appropriate when the observed correlation between successive residuals is large in absolute value.

4.7 Higher-Order AR Processes

Although AR(1) is a very flexible estimation method, there are many other stochastic processes that could reasonably be used to describe the evolution of error terms over time.

The AR(1) process is actually a special case of the **pth-order autoregressive**, or **AR(p)**, process

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \cdots + \rho_0 u_{t-p} + e_t \tag{131}$$

$$e_t \sim IID(0, \omega^2),$$

in which u$_t$ depends on up to p lagged values of itself, as well as on e_t. The AR(p) process can be expressed more compactly as

$$(1 - \rho_1 L - \rho_2 L^2 - \cdots - \rho_p L^p) u_t = e_t \tag{132}$$

$$e_t \sim IID(0, \omega^2),$$

where L denotes the **lag operator**. The lag operator L has the property that when L multiplies anything with a time subscript, this subscript is lagged one period. Thus,

$$L u_t = u_{t-1}, \tag{133}$$

$$L^2 u_t = u_{t-2}, \tag{134}$$

$$L^p u_t = u_{t-p}, \tag{135}$$

and so on. The expression in parentheses in Equation 132 is a polynomial in the lag operator L, with coefficients 1 and $-\rho_1, \cdots, -\rho_p$. If we define A(L, ρ) as begin equal to this polynomial, ρ representing the vector $\left[\rho_1 \vdots \rho_2 \vdots \cdots \vdots \rho_p\right]$, we can write Equation 132 even more compactly as

$$A(L, \rho)u_t = \epsilon_t, \tag{136}$$

$$e_t \sim IID(0, \omega^2).$$

For the same reasons that we wish to impose the condition $|\rho_1| < 1$ on AR(1) processes so as to ensure that they are stationary, we would like to impose stationarity conditions on general AR(p) processes. The stationarity condition for such processes may be expressed in several ways; one of them is that all the roots of the polynomial equation in z,

$$A(z, \rho) = 1 - \rho z - \rho_2 z^2 - \cdots - \rho_p z^p = 0 \tag{137}$$

must lie **outside the unit circle**, which simply means that all of the roots of Equation 137 must be greater than 1 in absolute value. This condition can lead to quite complicated restrictions on ρ in general AR(p) processes.

It rarely makes sense to specify a high-order AR(p) process—i.e., one with p a large number. The AR(2) process is much more flexible, but also much more complicated, than the AR(1) process; it is often all that is needed when AR(1) is too restrictive. The additional complexity of the AR(2) process is easily seen. For example, the variance of u_t, assuming stationarity, is

$$\sigma^2 = \frac{1 - \rho_2}{1 + \rho_2} \times \frac{\omega^2}{(1 - \rho_2)^2 - \rho_1^2}, \tag{138}$$

which is substantially more complicated than the corresponding expression for AR(1)..

Stationarity, in the case of AR(2), now requires that three conditions hold:

$$\rho_1 + \rho_2 < 1, \tag{139}$$

$$\rho_2 - \rho_1 < 1, \tag{140}$$

$$\rho_2 > -1. \tag{141}$$

These three conditions define a **stationarity triangle**. This triangle has vertices at (-2, -1), (2, -1) and (0,1). Provided that the point (ρ_1, ρ_2) lies within the triangle, the AR(2) process will be stationary.

Simple autoregressive processes of order higher than 2 arise quite frequently with time-series data that exhibit seasonal variation. It is not uncommon, for example, for error terms in models estimated using quarterly data apparently to follow the **simple AR(4) process**

$$u_t = \rho_4 u_{t-4} + \epsilon_t, \tag{142}$$

$$\epsilon \sim IID(0, \omega^2),$$

in which the error term in period t depends on the error in the same quarter of the previous year, but not on any intervening error terms. Another possibility is that the error terms my appear to follow a combined first- and fourth-order AR process

$$(1 - \rho_1 L)(1 - \rho_4 L^4)u_t = \epsilon_t, \tag{143}$$

$$\epsilon_t \sim IID(0, \omega^2).$$

If we multiply out the polynomial on the left-hand side, we obtain

$$(1 - \rho_1 L - \rho_4 L^4 + \rho_1 \rho_4 L^5)u_t = e_t, \tag{144}$$

$$\epsilon_t \sim IID(0, \omega^2).$$

This is a restricted special case of an AR(5) process, but with only two parameters instead of five to estimate. Various ways of modeling seasonality, including **seasonal AR processes**, will be discussed later.

Estimating a regression model with errors that follow an AR(p) process is not fundamentally different form estimating the same model with errors that follow an AR(1) process. Thus if, for example, we wish to estimate the model

$$y_t = x_t\beta + u_t, \tag{145}$$

where,

$$(1-\rho_1 L)(1-\rho_4 L^4)u_t = \epsilon_t,$$

$$\epsilon_t \sim IID(0, \omega^2),$$

we simply have to transform it into the model

$$y_t = x_t\beta + \rho_1(y_{t-1} - x_{t-1}\beta) + \rho_4(y_{t-4} - x_{t-4}\beta) - \rho_1\rho_4(y_{t-5} - x_{t-5}\beta) + \epsilon_t, \tag{146}$$

$$\epsilon_t \sim IID(0, \omega^2).$$

In order to estimate Equation 146 we have to drop the first *five* observations, and use nonlinear least squares. Dropping the first five observations may make us uncomfortable, especially if the sample size is modest, but it is certainly valid asymptotically.

4.8 ARMA and ARIMA Models

We may specify a model that combines both autoregressive and moving average components. The result is the so-called ARMA(p, q) model,

$$A(L,\rho)u_t = B(L,\alpha)\epsilon_t, \quad \epsilon_t \sim IID(0,\omega^2). \tag{147}$$

The left-hand side of Equation 147 looks like the AR(p) model, and the right-hand side looks like the MA(q) model. The advantage of ARMA models is that a relatively parsimonious model, such as ARMA(1,1) or ARMA(2,1), can often provide a representation of a time series which is as good as that obtained from a much less parsimonious AR or MA model.

ARIMA models are simply ARMA models applied to data that have been differenced some integer number of times, say d. Thus the **ARIMA(p, d, q)** model is

$$A(L,\rho)(1-L)^d u_t = B(L,\alpha)\epsilon_t, \quad \epsilon \sim IID(0,\omega^2). \tag{148}$$

When $d = 0$, this collapses to a standard ARMA(p,q) model. The I in ARIMA means integrated, since an integrated series has to be differenced to achieve stationarity. Differencing is often used to induce stationarity in time series that would otherwise be nonstationary.

4.9 Durbin Watson Statistic I

The standard Durbin Watson Statistic is designed to find first-order serial correlation, but only for linear models without lagged dependent variables and with error terms that are assumed to be normally distributed. This is the **d statistic** proposed by Durbin and Watson (1950, 1951) and commonly referred to as the DW statistic. It is defined as

$$d = \frac{\sum_{t=2}^{n}(\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^{n}\hat{u}_t^2} \tag{149}$$

where $\hat{u}_t$ is the t^{th} residual form OLS estimation of the regression that is being tested for possible first-order serial correlation. This regression may be linear or nonlinear, although finite-sample results depend on linearity.

The numerator of the d statistic is approximately equal to

$$2\left(\sum_{t=2}^{n}\hat{u}_t^2 - \sum_{t=2}^{n}\hat{u}_t\hat{u}_{t-1}\right). \tag{150}$$

Thus the d statistic itself is approximately equal to $2 - 2\hat{\rho}$, where $\hat{\rho}$ is the estimate of ρ obtained by regressing $\hat{u}_t$ on $\hat{u}_{t-1}$:

$$\hat{\rho} = \frac{\sum_{t=2}^{n}\hat{u}_t\hat{u}_{t-1}}{\sum_{t=2}^{n}\hat{u}_{t-1}^2} \tag{151}$$

These results are only approximations because all of the equations treat the first and last observations differently. Any effects of those observations must, however, vanish asymptotically. Thus it is clear that in samples of reasonable size the d statistic must vary between 0 and 4, and that a value of 2 corresponds to the complete absence of serial correlation. Values of the d statistic less than 2 correspond to $\hat{\rho} > 0$, while values greater than 2 correspond to $\hat{\rho} < 0$.

The DW d statistic is not valid, even asymptotically, when X includes a lagged dependent variable. This can be seen by observing that the d statistic is asymptotically equivalent to the t statistic on the estimation of ρ in the regression

$$\tilde{u} = \rho\tilde{u}_{t-1} + \text{residuals}. \tag{152}$$

This regression is similar to the general test for serial correlation:

$$\tilde{u} = \tilde{X}\beta + \rho\tilde{u}_{t-1} + \text{residuals}, \tag{153}$$

where $\tilde{u}$ denotes the vector of least squares residuals, $\tilde{X}$ denotes the matrix of derivatives of the regression function $X\beta$ with respect to β, and $\tilde{u}_{t-1}$ denotes the vector of the least squares residuals lagged one period.

When the original equation of interest is linear, Equation 153 is simple to setup. For then, $\tilde{X}$ is just X.

The test statistic of interest is the ordinary t statistic for $\rho = 0$.

This general test proposed in Equation 153 works even if there is a lagged dependent variable. This is because the lagged dependent variable is, in the linear case, included in $\tilde{X}$. This is just not the case in Equation 152. And since X contains lagged values of the dependent variable it will be, by definition, correlated with $\tilde{u}$. And if this is not taken into consideration, we may find serial correlation when there is none.

But the d statistic may still be informative even if there is a lagged dependent variable. If its value is such that we could reject the null hypothesis for no serial correlation, then the correct statistic based on Equation 153 would certainly allow us to do so.