Chapter Ten

Basic Regression Analysis with Time Series Data

In this chapter, we begin to study the properties of OLS for estimating linear regression models using time series data. In Section 10.1, we discuss some conceptual differences between time series and cross-sectional data. Section 10.2 provides some examples of time series regressions that are often estimated in the empirical social sciences. We then turn our attention to the finite sample properties of the OLS estimators and state the Gauss-Markov assumptions and the classical linear model assumptions for time series regression. While these assumptions have features in common with those for the cross-sectional case, they also have some significant differences that we will need to highlight.

In addition, we return to some issues that we treated in regression with cross-sectional data, such as how to use and interpret the logarithmic functional form and dummy variables. The important topics of how to incorporate trends and account for seasonality in multiple regression are taken up in Section 10.5.

10.1 THE NATURE OF TIME SERIES DATA

An obvious characteristic of time series data which distinguishes it from cross-sectional data is that a time series data set comes with a temporal ordering. For example, in Chapter 1, we briefly discussed a time series data set on employment, the minimum wage, and other economic variables for Puerto Rico. In this data set, we must know that the data for 1970 immediately precede the data for 1971. For analyzing time series data in the social sciences, we must recognize that the past can effect the future, but not vice versa (unlike in the Star Trek universe). To emphasize the proper ordering of time series data, Table 10.1 gives a partial listing of the data on U.S. inflation and unemployment rates in PHILLIPS.RAW.

Another difference between cross-sectional and time series data is more subtle. In Chapters 3 and 4, we studied statistical properties of the OLS estimators based on the notion that samples were randomly drawn from the appropriate population. Understanding why cross-sectional data should be viewed as random outcomes is fairly straightforward: a different sample drawn from the population will generally yield different values of the independent and dependent variables (such as education, experience, wage, and so on). Therefore, the OLS estimates computed from different random samples will generally differ, and this is why we consider the OLS estimators to be random variables.

Table 10.1

Partial Listing of Data on U.S. Inflation and Unemployment Rates, 1948–1996

Year	Inflation	Unemployment
1948	8.1	3.8
1949	−1.2	5.9
1950	1.3	5.3
1951	7.9	3.3
. . .	. . .	. . .
1994	2.6	6.1
1995	2.8	5.6
1996	3.0	5.4

How should we think about randomness in time series data? Certainly, economic time series satisfy the intuitive requirements for being outcomes of random variables. For example, today we do not know what the Dow Jones Industrial Average will be at its close at the end of the next trading day. We do not know what the annual growth in output will be in Canada during the coming year. Since the outcomes of these variables are not foreknown, they should clearly be viewed as random variables.

Formally, a sequence of random variables indexed by time is called a **stochastic process** or a **time series process**. ("Stochastic" is a synonym for random.) When we collect a time series data set, we obtain one possible outcome, or *realization*, of the stochastic process. We can only see a single realization, because we cannot go back in time and start the process over again. (This is analogous to cross-sectional analysis where we can collect only one random sample.) However, if certain conditions in history had been different, we would generally obtain a different realization for the stochastic process, and this is why we think of time series data as the outcome of random variables. The set of all possible realizations of a time series process plays the role of the population in cross-sectional analysis.

10.2 EXAMPLES OF TIME SERIES REGRESSION MODELS

In this section, we discuss two examples of time series models that have been useful in empirical time series analysis and that are easily estimated by ordinary least squares. We will study additional models in Chapter 11.

Static Models

Suppose that we have time series data available on two variables, say y and z, where y_t and z_t are dated contemporaneously. A **static model** relating y to z is

$$y_t = \beta_0 + \beta_1 z_t + u_t, t = 1,2, \ldots, n. \quad \textbf{(10.1)}$$

The name "static model" comes from the fact that we are modeling a contemporaneous relationship between y and z. Usually, a static model is postulated when a change in z at time t is believed to have an immediate effect on y: $\Delta y_t = \beta_1 \Delta z_t$, when $\Delta u_t = 0$. Static regression models are also used when we are interested in knowing the tradeoff between y and z.

An example of a static model is the *static Phillips curve*, given by

$$inf_t = \beta_0 + \beta_1 unem_t + u_t, \quad \textbf{(10.2)}$$

where inf_t is the annual inflation rate and $unem_t$ is the unemployment rate. This form of the Phillips curve assumes a constant *natural rate of unemployment* and constant inflationary expectations, and it can be used to study the contemporaneous tradeoff between them. [See, for example, Mankiw (1994, Section 11.2).]

Naturally, we can have several explanatory variables in a static regression model. Let $mrdrte_t$ denote the murders per 10,000 people in a particular city during year t, let $convrte_t$ denote the murder conviction rate, let $unem_t$ be the local unemployment rate, and let $yngmle_t$ be the fraction of the population consisting of males between the ages of 18 and 25. Then, a static multiple regression model explaining murder rates is

$$mrdrte_t = \beta_0 + \beta_1 convrte_t + \beta_2 unem_t + \beta_3 yngmle_t + u_t. \quad \textbf{(10.3)}$$

Using a model such as this, we can hope to estimate, for example, the ceteris paribus effect of an increase in the conviction rate on criminal activity.

Finite Distributed Lag Models

In a **finite distributed lag (FDL) model**, we allow one or more variables to affect y with a lag. For example, for annual observations, consider the model

$$gfr_t = \alpha_0 + \delta_0 pe_t + \delta_1 pe_{t-1} + \delta_2 pe_{t-2} + u_t, \quad \textbf{(10.4)}$$

where gfr_t is the general fertility rate (children born per 1,000 women of childbearing age) and pe_t is the real dollar value of the personal tax exemption. The idea is to see whether, in the aggregate, the decision to have children is linked to the tax value of having a child. Equation (10.4) recognizes that, for both biological and behavioral reasons, decisions to have children would not immediately result from changes in the personal exemption.

Equation (10.4) is an example of the model

$$y_t = \alpha_0 + \delta_0 z_t + \delta_1 z_{t-1} + \delta_2 z_{t-2} + u_t, \quad \textbf{(10.5)}$$

which is an FDL *of order two*. To interpret the coefficients in (10.5), suppose that z is a constant, equal to c, in all time periods before time t. At time t, z increases by one unit to $c + 1$ and then reverts to its previous level at time $t + 1$. (That is, the increase in z is temporary.) More precisely,

$$\ldots, z_{t-2} = c, z_{t-1} = c, z_t = c + 1, z_{t+1} = c, z_{t+2} = c, \ldots.$$

To focus on the ceteris paribus effect of z on y, we set the error term in each time period to zero. Then,

$$y_{t-1} = \alpha_0 + \delta_0 c + \delta_1 c + \delta_2 c,$$
$$y_t = \alpha_0 + \delta_0(c + 1) + \delta_1 c + \delta_2 c,$$
$$y_{t+1} = \alpha_0 + \delta_0 c + \delta_1(c + 1) + \delta_2 c,$$
$$y_{t+2} = \alpha_0 + \delta_0 c + \delta_1 c + \delta_2(c + 1),$$
$$y_{t+3} = \alpha_0 + \delta_0 c + \delta_1 c + \delta_2 c,$$

and so on. From the first two equations, $y_t - y_{t-1} = \delta_0$, which shows that δ_0 is the immediate change in y due to the one-unit increase in z at time t. δ_0 is usually called the **impact propensity** or **impact multiplier**.

Similarly, $\delta_1 = y_{t+1} - y_{t-1}$ is the change in y one period after the temporary change, and $\delta_2 = y_{t+2} - y_{t-1}$ is the change in y two periods after the change. At time $t + 3$, y has reverted back to its initial level: $y_{t+3} = y_{t-1}$. This is because we have assumed that only two lags of z appear in (10.5). When we graph the δ_j as a function of j, we obtain the **lag distribution**, which summarizes the dynamic effect that a temporary increase in z has on y. A possible lag distribution for the FDL of order two is given in Figure 10.1. (Of course, we would never know the parameters δ_j; instead, we will estimate the δ_j and then plot the estimated lag distribution.)

The lag distribution in Figure 10.1 implies that the largest effect is at the first lag. The lag distribution has a useful interpretation. If we standardize the initial value of y at $y_{t-1} = 0$, the lag distribution traces out all subsequent values of y due to a one-unit, temporary increase in z.

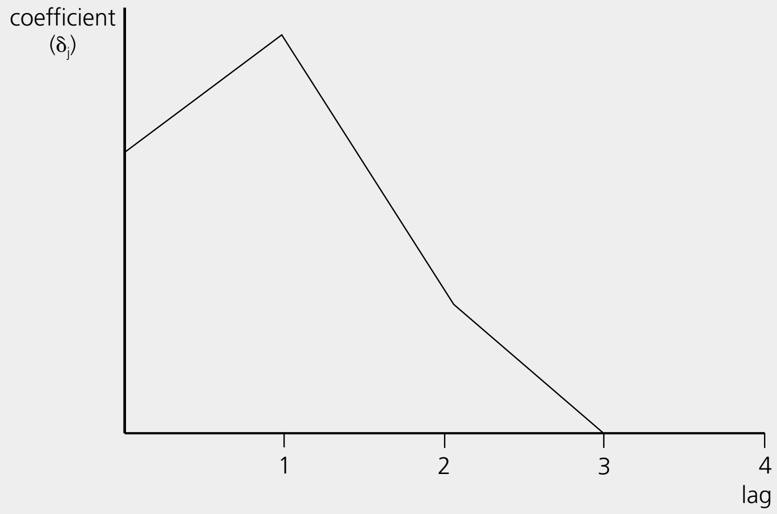
We are also interested in the change in y due to a *permanent* increase in z. Before time t, z equals the constant c. At time t, z increases permanently to $c + 1$: $z_s = c$, $s < t$ and $z_s = c + 1$, $s \geq t$. Again, setting the errors to zero, we have

$$y_{t-1} = \alpha_0 + \delta_0 c + \delta_1 c + \delta_2 c,$$
$$y_t = \alpha_0 + \delta_0(c + 1) + \delta_1 c + \delta_2 c,$$
$$y_{t+1} = \alpha_0 + \delta_0(c + 1) + \delta_1(c + 1) + \delta_2 c,$$
$$y_{t+2} = \alpha_0 + \delta_0(c + 1) + \delta_1(c + 1) + \delta_2(c + 1),$$

and so on. With the permanent increase in z, after one period, y has increased by $\delta_0 + \delta_1$, and after two periods, y has increased by $\delta_0 + \delta_1 + \delta_2$. There are no further changes in y after two periods. This shows that the sum of the coefficients on current and lagged z, $\delta_0 + \delta_1 + \delta_2$, is the *long-run* change in y given a permanent increase in z and is called the **long-run propensity** (**LRP**) or **long-run multiplier**. The LRP is often of interest in distributed lag models.

Figure 10.1

A lag distribution with two nonzero lags. The maximum effect is at the first lag.



As an example, in equation (10.4), δ_0 measures the immediate change in fertility due to a one-dollar increase in *pe*. As we mentioned earlier, there are reasons to believe that δ_0 is small, if not zero. But δ_1 or δ_2, or both, might be positive. If *pe permanently* increases by one dollar, then, after two years, *gfr* will have changed by $\delta_0 + \delta_1 + \delta_2$. This model assumes that there are no further changes after two years. Whether or not this is actually the case is an empirical matter.

A finite distributed lag model of order q is written as

$$y_t = \alpha_0 + \delta_0 z_t + \delta_1 z_{t-1} + \ldots + \delta_q z_{t-q} + u_t. \quad \textbf{(10.6)}$$

This contains the static model as a special case by setting $\delta_1, \delta_2, \ldots, \delta_q$ equal to zero. Sometimes, a primary purpose for estimating a distributed lag model is to test whether z has a lagged effect on y. The impact propensity is always the coefficient on the contemporaneous z, δ_0. Occasionally, we omit z_t from (10.6), in which case the impact propensity is zero. The lag distribution is again the δ_j graphed as a function of j. The long-run propensity is the sum of all coefficients on the variables z_{t-j}:

$$\text{LRP} = \delta_0 + \delta_1 + \ldots + \delta_q. \quad \textbf{(10.7)}$$

Because of the often substantial correlation in z at different lags—that is, due to multicollinearity in (10.6)—it can be difficult to obtain precise estimates of the individual δ_j. Interestingly, even when the δ_j cannot be precisely estimated, we can often get good estimates of the LRP. We will see an example later.

QUESTION 10.1

In an equation for annual data, suppose that

$$int_t = 1.6 + .48\ inf_t - .15\ inf_{t-1} + .32\ inf_{t-2} + u_t,$$

where *int* is an interest rate and *inf* is the inflation rate, what are the impact and long-run propensities?

We can have more than one explanatory variable appearing with lags, or we can add contemporaneous variables to an FDL model. For example, the average education level for women of childbearing age could be added to (10.4), which allows us to account for changing education levels for women.

A Convention About the Time Index

When models have lagged explanatory variables (and, as we will see in the next chapter, models with lagged y), confusion can arise concerning the treatment of initial observations. For example, if in (10.5), we assume that the equation holds, starting at $t = 1$, then the explanatory variables for the first time period are z_1, z_0, and z_{-1}. Our convention will be that these are the initial values in our sample, so that we can always start the time index at $t = 1$. In practice, this is not very important because regression packages automatically keep track of the observations available for estimating models with lags. But for this and the next few chapters, we need some convention concerning the first time period being represented by the regression equation.

10.3 FINITE SAMPLE PROPERTIES OF OLS UNDER CLASSICAL ASSUMPTIONS

In this section, we give a complete listing of the finite sample, or small sample, properties of OLS under standard assumptions. We pay particular attention to how the assumptions must be altered from our cross-sectional analysis to cover time series regressions.

Unbiasedness of OLS

The first assumption simply states that the time series process follows a model which is linear in its parameters.

ASSUMPTION TS.1 (LINEAR IN PARAMETERS)

The stochastic process $\{(x_{t1}, x_{t2}, \ldots, x_{tk}, y_t): t = 1,2, \ldots, n\}$ follows the linear model

$$y_t = \beta_0 + \beta_1 x_{t1} + \ldots + \beta_k x_{tk} + u_t, \qquad \textbf{(10.8)}$$

where $\{u_t: t = 1,2, \ldots, n\}$ is the sequence of errors or disturbances. Here, n is the number of observations (time periods).

Table 10.2

Example of **X** for the Explanatory Variables in Equation (10.3)

t	*convrte*	*unem*	*yngmle*
1	.46	.074	.12
2	.42	.071	.12
3	.42	.063	.11
4	.47	.062	.09
5	.48	.060	.10
6	.50	.059	.11
7	.55	.058	.12
8	.56	.059	.13

In the notation x_{tj}, t denotes the time period, and j is, as usual, a label to indicate one of the k explanatory variables. The terminology used in cross-sectional regression applies here: y_t is the dependent variable, explained variable, or regressand; the x_{tj} are the independent variables, explanatory variables, or regressors.

We should think of Assumption TS.1 as being essentially the same as Assumption MLR.1 (the first cross-sectional assumption), but we are now specifying a linear model for time series data. The examples covered in Section 10.2 can be cast in the form of (10.8) by appropriately defining x_{tj}. For example, equation (10.5) is obtained by setting $x_{t1} = z_t$, $x_{t2} = z_{t-1}$, and $x_{t3} = z_{t-2}$.

In order to state and discuss several of the remaining assumptions, we let $\mathbf{x}_t = (x_{t1}, x_{t2}, \ldots, x_{tk})$ denote the set all independent variables in the equation at time t. Further, $\mathbf{X}$ denotes the collection of all independent variables for all time periods. It is useful to think of $\mathbf{X}$ as being an array, with n rows and k columns. This reflects how time series data are stored in econometric software packages: the t^{th} row of $\mathbf{X}$ is $\mathbf{x}_t$, consisting of all independent variables for time period t. Therefore, the first row of $\mathbf{X}$ corresponds to $t = 1$, the second row to $t = 2$, and the last row to $t = n$. An example is given in Table 10.2, using $n = 8$ and the explanatory variables in equation (10.3).

The next assumption is the time series analog of Assumption MLR.3, and it also drops the assumption of random sampling in Assumption MLR.2.

ASSUMPTION TS.2 (ZERO CONDITIONAL MEAN)

For each t, the expected value of the error u_t, given the explanatory variables for *all* time periods, is zero. Mathematically,

$$\mathrm{E}(u_t|\boldsymbol{X}) = 0, t = 1,2, \ldots, n. \tag{10.9}$$

This is a crucial assumption, and we need to have an intuitive grasp of its meaning. As in the cross-sectional case, it is easiest to view this assumption in terms of uncorrelatedness: Assumption TS.2 implies that the error at time t, u_t, is uncorrelated with each explanatory variable in *every* time period. The fact that this is stated in terms of the conditional expectation means that we must also correctly specify the functional relationship between y_t and the explanatory variables. If u_t is independent of $\boldsymbol{X}$ and $\mathrm{E}(u_t) = 0$, then Assumption TS.2 automatically holds.

Given the cross-sectional analysis from Chapter 3, it is not surprising that we require u_t to be uncorrelated with the explanatory variables also dated at time t: in conditional mean terms,

$$\mathrm{E}(u_t|x_{t1}, \ldots, x_{tk}) = \mathrm{E}(u_t|\boldsymbol{x}_t) = 0. \tag{10.10}$$

When (10.10) holds, we say that the x_{tj} are **contemporaneously exogenous**. Equation (10.10) implies that u_t and the explanatory variables are contemporaneously uncorrelated: $\mathrm{Corr}(x_{tj},u_t) = 0$, for all j.

Assumption TS.2 requires more than contemporaneous exogeneity: u_t must be uncorrelated with x_{sj}, even when $s \neq t$. This is a strong sense in which the explanatory variables must be exogenous, and when TS.2 holds, we say that the explanatory variables are **strictly exogenous**. In Chapter 11, we will demonstrate that (10.10) is sufficient for proving consistency of the OLS estimator. But to show that OLS is unbiased, we need the strict exogeneity assumption.

In the cross-sectional case, we did not explicitly state how the error term for, say, person i, u_i, is related to the explanatory variables for *other* people in the sample. The reason this was unnecessary is that, with random sampling (Assumption MLR.2), u_i is *automatically* independent of the explanatory variables for observations other than i. In a time series context, random sampling is almost never appropriate, so we must explicitly assume that the expected value of u_t is not related to the explanatory variables in any time periods.

It is important to see that Assumption TS.2 puts no restriction on correlation in the independent variables or in the u_t across time. Assumption TS.2 only says that the average value of u_t is unrelated to the independent variables in all time periods.

Anything that causes the unobservables at time t to be correlated with any of the explanatory variables in any time period causes Assumption TS.2 to fail. Two leading candidates for failure are omitted variables and measurement error in some of the regressors. But, the strict exogeneity assumption can also fail for other, less obvious reasons. In the simple static regression model

$$y_t = \beta_0 + \beta_1 z_t + u_t,$$

Assumption TS.2 requires not only that u_t and z_t are uncorrelated, but that u_t is also uncorrelated with past and future values of z. This has two implications. First, z can have no lagged effect on y. If z does have a lagged effect on y, then we should estimate a distributed lag model. A more subtle point is that strict exogeneity excludes the possibility that changes in the error term today can cause future changes in z. This effectively rules out feedback from y on future values of z. For example, consider a simple static model to explain a city's murder rate in terms of police officers per capita:

$$mrdrte_t = \beta_0 + \beta_1 polpc_t + u_t.$$

It may be reasonable to assume that u_t is uncorrelated with $polpc_t$ and even with past values of $polpc_t$; for the sake of argument, assume this is the case. But suppose that the city adjusts the size of its police force based on past values of the murder rate. This means that, say, $polpc_{t+1}$ might be correlated with u_t (since a higher u_t leads to a higher $mrdrte_t$). If this is the case, Assumption TS.2 is generally violated.

There are similar considerations in distributed lag models. Usually we do not worry that u_t might be correlated with past z because we are controlling for past z in the model. But feedback from u to future z is always an issue.

Explanatory variables that are strictly exogenous cannot react to what has happened to y in the past. A factor such as the amount of rainfall in an agricultural production function satisfies this requirement: rainfall in any future year is not influenced by the output during the current or past years. But something like the amount of labor input might not be strictly exogenous, as it is chosen by the farmer, and the farmer may adjust the amount of labor based on last year's yield. Policy variables, such as growth in the money supply, expenditures on welfare, highway speed limits are often influenced by what has happened to the outcome variable in the past. In the social sciences, many explanatory variables may very well violate the strict exogeneity assumption.

Even though Assumption TS.2 can be unrealistic, we begin with it in order to conclude that the OLS estimators are unbiased. Most treatments of static and finite distributed lag models assume TS.2 by making the stronger assumption that the explanatory variables are nonrandom, or fixed in repeated samples. The nonrandomness assumption is obviously false for time series observations; Assumption TS.2 has the advantage of being more realistic about the random nature of the x_{tj}, while it isolates the necessary assumption about how u_t and the explanatory variables are related in order for OLS to be unbiased.

The last assumption needed for unbiasedness of OLS is the standard no perfect collinearity assumption.

ASSUMPTION TS.3 (NO PERFECT COLLINEARITY)

In the sample (and therefore in the underlying time series process), no independent variable is constant or a perfect linear combination of the others.

We discussed this assumption at length in the context of cross-sectional data in Chapter 3. The issues are essentially the same with time series data. Remember, Assumption TS.3 does allow the explanatory variables to be correlated, but it rules out *perfect* correlation in the sample.

THEOREM 10.1 (UNBIASEDNESS OF OLS)

Under Assumptions TS.1, TS.2, and TS.3, the OLS estimators are unbiased conditional on $\mathbf{X}$, and therefore unconditionally as well: $E(\hat{\beta}_j) = \beta_j, j = 0,1, \dots, k$.

QUESTION 10.2

In the FDL model $y_t = \alpha_0 + \delta_0 z_t + \delta_1 z_{t-1} + u_t$, what do we need to assume about the sequence $\{z_0, z_1, \dots, z_n\}$ in order for Assumption TS.3 to hold?

The proof of this theorem is essentially the same as that for Theorem 3.1 in Chapter 3, and so we omit it. When comparing Theorem 10.1 to Theorem 3.1, we have been able to drop the random sampling assumption by assuming that, for each t, u_t has zero mean given the explanatory variables at all time periods. If this assumption does not hold, OLS cannot be shown to be unbiased.

The analysis of omitted variables bias, which we covered in Section 3.3, is essentially the same in the time series case. In particular, Table 3.2 and the discussion surrounding it can be used as before to determine the directions of bias due to omitted variables.

The Variances of the OLS Estimators and the Gauss-Markov Theorem

We need to add two assumptions to round out the Gauss-Markov assumptions for time series regressions. The first one is familiar from cross-sectional analysis.

ASSUMPTION TS.4 (HOMOSKEDASTICITY)

Conditional on $\boldsymbol{X}$, the variance of u_t is the same for all t: $\text{Var}(u_t|\boldsymbol{X}) = \text{Var}(u_t) = \sigma^2$, $t = 1,2,\dots,n$.

This assumption means that $\text{Var}(u_t|\boldsymbol{X})$ cannot depend on $\boldsymbol{X}$—it is sufficient that u_t and $\boldsymbol{X}$ are independent—*and* that $\text{Var}(u_t)$ must be constant over time. When TS.4 does not hold, we say that the errors are *heteroskedastic*, just as in the cross-sectional case. For example, consider an equation for determining three-month, T-bill rates ($i3_t$) based on the inflation rate (inf_t) and the federal deficit as a percentage of gross domestic product (def_t):

$$i3_t = \beta_0 + \beta_1 inf_t + \beta_2 def_t + u_t. \tag{10.11}$$

Among other things, Assumption TS.4 requires that the unobservables affecting interest rates have a constant variance over time. Since policy regime changes are known to affect the variability of interest rates, this assumption might very well be false. Further, it could be that the variability in interest rates depends on the level of inflation or relative size of the deficit. This would also violate the homoskedasticity assumption.

When $\text{Var}(u_t|\boldsymbol{X})$ does depend on $\boldsymbol{X}$, it often depends on the explanatory variables at time t, $\boldsymbol{x}_t$. In Chapter 12, we will see that the tests for heteroskedasticity from Chapter 8 can also be used for time series regressions, at least under certain assumptions.

The final Gauss-Markov assumption for time series analysis is new.

ASSUMPTION TS.5 (NO SERIAL CORRELATION)

Conditional on $\boldsymbol{X}$, the errors in two different time periods are uncorrelated: $\text{Corr}(u_t,u_s|\boldsymbol{X}) = 0$, for all $t \neq s$.

The easiest way to think of this assumption is to ignore the conditioning on $\boldsymbol{X}$. Then, Assumption TS.5 is simply

$$\text{Corr}(u_t, u_s) = 0, \text{ for all } t \neq s. \quad \textbf{(10.12)}$$

(This is how the no serial correlation assumption is stated when $\boldsymbol{X}$ is treated as nonrandom.) When considering whether Assumption TS.5 is likely to hold, we focus on equation (10.12) because of its simple interpretation.

When (10.12) is false, we say that the errors in (10.8) suffer from **serial correlation**, or **autocorrelation**, because they are correlated across time. Consider the case of errors from adjacent time periods. Suppose that, when $u_{t-1} > 0$ then, on average, the error in the next time period, u_t, is also positive. Then $\text{Corr}(u_t, u_{t-1}) > 0$, and the errors suffer from serial correlation. In equation (10.11) this means that, if interest rates are unexpectedly high for this period, then they are likely to be above average (for the given levels of inflation and deficits) for the next period. This turns out to be a reasonable characterization for the error terms in many time series applications, which we will see in Chapter 12. For now, we assume TS.5.

Importantly, Assumption TS.5 assumes nothing about temporal correlation in the *independent* variables. For example, in equation (10.11), inf_t is almost certainly correlated across time. But this has nothing to do with whether TS.5 holds.

A natural question that arises is: In Chapters 3 and 4, why did we not assume that the errors for different cross-sectional observations are uncorrelated? The answer comes from the random sampling assumption: under random sampling, u_i and u_h are independent for any two observations i and h. It can also be shown that this is true, conditional on all explanatory variables in the sample. Thus, for our purposes, serial correlation is only an issue in time series regressions.

Assumptions TS.1 through TS.5 are the appropriate Gauss-Markov assumptions for time series applications, but they have other uses as well. Sometimes, TS.1 through TS.5 are satisfied in cross-sectional applications, even when random sampling is not a reasonable assumption, such as when the cross-sectional units are large relative to the population. It is possible that correlation exists, say, across cities within a state, but as long as the errors are uncorrelated across those cities, Assumption TS.5 holds. But we are primarily interested in applying these assumptions to regression models with time series data.

THEOREM 10.2 (OLS SAMPLING VARIANCES)

Under the time series Gauss-Markov assumptions TS.1 through TS.5, the variance of $\hat{\beta}_j$, conditional on $\boldsymbol{X}$, is

$$\text{Var}(\hat{\beta}_j | \boldsymbol{X}) = \sigma^2 / [\text{SST}_j (1 - R_j^2)], \; j = 1, \ldots, k, \quad \textbf{(10.13)}$$

where SST_j is the total sum of squares of x_{tj} and R_j^2 is the R-squared from the regression of x_j on the other independent variables.

Equation (10.13) is the exact variance we derived in Chapter 3 under the cross-sectional Gauss-Markov assumptions. Since the proof is very similar to the one for Theorem 3.2, we omit it. The discussion from Chapter 3 about the factors causing large variances, including multicollinearity among the explanatory variables, applies immediately to the time series case.

The usual estimator of the error variance is also unbiased under Assumptions TS.1 through TS.5, and the Gauss-Markov theorem holds.

THEOREM 10.3 (UNBIASED ESTIMATION OF σ^2)

Under Assumptions TS.1 through TS.5, the estimator $\hat{\sigma}^2 = \text{SSR}/df$ is an unbiased estimator of σ^2, where $df = n - k - 1$.

THEOREM 10.4 (GAUSS-MARKOV THEOREM)

Under Assumptions TS.1 through TS.5, the OLS estimators are the best linear unbiased estimators conditional on $\boldsymbol{X}$.

QUESTION 10.3

In the FDL model $y_t = \alpha_0 + \delta_0 z_t + \delta_1 z_{t-1} + u_t$, explain the nature of any multicollinearity in the explanatory variables.

The bottom line here is that OLS has the same desirable finite sample properties under TS.1 through TS.5 that it has under MLR.1 through MLR.5.

Inference Under the Classical Linear Model Assumptions

In order to use the usual OLS standard errors, *t* statistics, and *F* statistics, we need to add a final assumption that is analogous to the normality assumption we used for cross-sectional analysis.

ASSUMPTION TS.6 (NORMALITY)

The errors u_t are independent of $\boldsymbol{X}$ and are independently and identically distributed as Normal$(0,\sigma^2)$.

Assumption TS.6 implies TS.3, TS.4, and TS.5, but it is stronger because of the independence and normality assumptions.

THEOREM 10.5 (NORMAL SAMPLING DISTRIBUTIONS)

Under Assumptions TS.1 through TS.6, the CLM assumptions for time series, the OLS estimators are normally distributed, conditional on $\boldsymbol{X}$. Further, under the null hypothesis, each *t* statistic has a *t* distribution, and each *F* statistic has an *F* distribution. The usual construction of confidence intervals is also valid.

The implications of Theorem 10.5 are of utmost importance. It implies that, when Assumptions TS.1 through TS.6 hold, everything we have learned about estimation and inference for cross-sectional regressions applies directly to time series regressions. Thus, t statistics can be used for testing statistical significance of individual explanatory variables, and F statistics can be used to test for joint significance.

Just as in the cross-sectional case, the usual inference procedures are only as good as the underlying assumptions. The classical linear model assumptions for time series data are much more restrictive than those for the cross-sectional data—in particular, the strict exogeneity and no serial correlation assumptions can be unrealistic. Nevertheless, the CLM framework is a good starting point for many applications.

EXAMPLE 10.1
(Static Phillips Curve)

To determine whether there is a tradeoff, on average, between unemployment and inflation, we can test H_0: $\beta_1 = 0$ against H_0: $\beta_1 < 0$ in equation (10.2). If the classical linear model assumptions hold, we can use the usual OLS t statistic. Using annual data for the United States in PHILLIPS.RAW, for the years 1948 through 1996, we obtain

$$\begin{gathered} \widehat{inf}_t = 1.42 + .468\ unem_t \\ (1.72) \quad (.289) \\ n = 49,\ R^2 = .053,\ \bar{R}^2 = .033. \end{gathered} \tag{10.14}$$

This equation does not suggest a tradeoff between *unem* and *inf*: $\hat{\beta}_1 > 0$. The t statistic for $\hat{\beta}_1$ is about 1.62, which gives a p-value against a two-sided alternative of about .11. Thus, if anything, there is a positive relationship between inflation and unemployment.

There are some problems with this analysis that we cannot address in detail now. In Chapter 12, we will see that the CLM assumptions do not hold. In addition, the static Phillips curve is probably not the best model for determining whether there is a short-run tradeoff between inflation and unemployment. Macroeconomists generally prefer the expectations augmented Phillips curve, a simple example of which is given in Chapter 11.

As a second example, we estimate equation (10.11) using anual data on the U.S. economy.

EXAMPLE 10.2
(Effects of Inflation and Deficits on Interest Rates)

The data in INTDEF.RAW come from the 1997 *Economic Report of the President* and span the years 1948 through 1996. The variable *i3* is the three-month T-bill rate, *inf* is the annual inflation rate based on the consumer price index (CPI), and *def* is the federal budget deficit as a percentage of GDP. The estimated equation is

$$\hat{i3}_t = 1.25 + .613\, inf_t + .700\, def_t$$
$$(0.44)\quad (.076)\qquad (.118)$$
$$n = 49,\ R^2 = .697,\ \bar{R}^2 = .683. \tag{10.15}$$

These estimates show that increases in inflation and the relative size of the deficit work together to increase short-term interest rates, both of which are expected from basic economics. For example, a ceteris paribus one percentage point increase in the inflation rate increases *i3* by .613 points. Both *inf* and *def* are very statistically significant, assuming, of course, that the CLM assumptions hold.

10.4 FUNCTIONAL FORM, DUMMY VARIABLES, AND INDEX NUMBERS

All of the functional forms we learned about in earlier chapters can be used in time series regressions. The most important of these is the natural logarithm: time series regressions with constant percentage effects appear often in applied work.

EXAMPLE 10.3

(Puerto Rican Employment and the Minimum Wage)

Annual data on the Puerto Rican employment rate, minimum wage, and other variables are used by Castillo-Freedman and Freedman (1992) to study the effects of the U.S. minimum wage on employment in Puerto Rico. A simplified version of their model is

$$\log(prepop_t) = \beta_0 + \beta_1\log(mincov_t) + \beta_2\log(usgnp_t) + u_t, \tag{10.16}$$

where $prepop_t$ is the employment rate in Puerto Rico during year *t* (ratio of those working to total population), $usgnp_t$ is real U.S. gross national product (in billions of dollars), and *mincov* measures the importance of the minimum wage relative to average wages. In particular, *mincov* = (*avgmin*/*avgwage*)·*avgcov*, where *avgmin* is the average minimum wage, *avgwage* is the average overall wage, and *avgcov* is the average coverage rate (the proportion of workers actually covered by the minimum wage law).

Using data for the years 1950 through 1987 gives

$$\log(\widehat{prepop}_t) = -1.05 - .154\,\log(mincov_t) - .012\,\log(usgnp_t)$$
$$(0.77)\quad (.065)\qquad\qquad (.089)$$
$$n = 38,\ R^2 = .661,\ \bar{R}^2 = .641. \tag{10.17}$$

The estimated elasticity of *prepop* with respect to *mincov* is −.154, and it is statistically significant with $t = -2.37$. Therefore, a higher minimum wage lowers the employment rate, something that classical economics predicts. The GNP variable is not statistically significant, but this changes when we account for a time trend in the next section.

We can use logarithmic functional forms in distributed lag models, too. For example, for quarterly data, suppose that money demand (M_t) and gross domestic product (GDP_t) are related by

$$\log(M_t) = \alpha_0 + \delta_0\log(GDP_t) + \delta_1\log(GDP_{t-1}) + \delta_2\log(GDP_{t-2}) + \delta_3\log(GDP_{t-3}) + \delta_4\log(GDP_{t-4}) + u_t.$$

The impact propensity in this equation, δ_0, is also called the **short-run elasticity**: it measures the immediate percentage change in money demand given a 1% increase in *GDP*. The long-run propensity, $\delta_0 + \delta_1 + \ldots + \delta_4$, is sometimes called the **long-run elasticity**: it measures the percentage increase in money demand after four quarters given a permanent 1% increase in *GDP*.

Binary or dummy independent variables are also quite useful in time series applications. Since the unit of observation is time, a dummy variable represents whether, in each time period, a certain event has occurred. For example, for annual data, we can indicate in each year whether a Democrat or a Republican is president of the United States by defining a variable $democ_t$, which is unity if the president is a Democrat, and zero otherwise. Or, in looking at the effects of capital punishment on murder rates in Texas, we can define a dummy variable for each year equal to one if Texas had capital punishment during that year, and zero otherwise.

Often dummy variables are used to isolate certain periods that may be systematically different from other periods covered by a data set.

EXAMPLE 10.4
(Effects of Personal Exemption on Fertility Rates)

The general fertility rate (*gfr*) is the number of children born to every 1,000 women of childbearing age. For the years 1913 through 1984, the equation,

$$gfr_t = \beta_0 + \beta_1 pe_t + \beta_2 ww2_t + \beta_3 pill_t + u_t,$$

explains *gfr* in terms of the average real dollar value of the personal tax exemption (*pe*) and two binary variables. The variable *ww2* takes on the value unity during the years 1941 through 1945, when the United States was involved in World War II. The variable *pill* is unity from 1963 on, when the birth control pill was made available for contraception.

Using the data in FERTIL3.RAW, which were taken from the article by Whittington, Alm, and Peters (1990), gives

$$\begin{aligned} \widehat{gfr}_t &= \underset{(3.21)}{98.68} + \underset{(.030)}{.083}\, pe_t - \underset{(7.46)}{24.24}\, ww2_t - \underset{(4.08)}{31.59}\, pill_t \\ n &= 72,\ R^2 = .473,\ \bar{R}^2 = .450. \end{aligned} \tag{10.18}$$

Each variable is statistically significant at the 1% level against a two-sided alternative. We see that the fertility rate was lower during World War II: given *pe*, there were about 24 fewer births for every 1,000 women of childbearing age, which is a large reduction. (From 1913 through 1984, *gfr* ranged from about 65 to 127.) Similarly, the fertility rate has been substantially lower since the introduction of the birth control pill.

The variable of economic interest is *pe*. The average *pe* over this time period is \$100.40, ranging from zero to \$243.83. The coefficient on *pe* implies that a 12-dollar increase in *pe* increases *gfr* by about one birth per 1,000 women of childbearing age. This effect is hardly trivial.

In Section 10.2, we noted that the fertility rate may react to changes in *pe* with a lag. Estimating a distributed lag model with two lags gives

$$\begin{aligned} \widehat{gfr}_t = {} & \underset{(3.28)}{95.87} + \underset{(.126)}{.073}\, pe_t - \underset{(.1557)}{.0058}\, pe_{t-1} + \underset{(.126)}{.034}\, pe_{t-2} \\ & - \underset{(10.73)}{22.13}\, ww2_t - \underset{(3.98)}{31.30}\, pill_t \\ & n = 70,\ R^2 = .499,\ \bar{R}^2 = .459. \end{aligned} \tag{10.19}$$

In this regression, we only have 70 observations because we lose two when we lag *pe* twice. The coefficients on the *pe* variables are estimated very imprecisely, and each one is individually insignificant. It turns out that there is substantial correlation between pe_t, pe_{t-1}, and pe_{t-2}, and this multicollinearity makes it difficult to estimate the effect at each lag. However, pe_t, pe_{t-1}, and pe_{t-2} are jointly significant: the *F* statistic has a *p*-value = .012. Thus, *pe* does have an effect on *gfr* [as we already saw in (10.18)], but we do not have good enough estimates to determine whether it is contemporaneous or with a one- or two-year lag (or some of each). Actually, pe_{t-1} and pe_{t-2} are jointly insignificant in this equation (*p*-value = .95), so at this point, we would be justified in using the static model. But for illustrative purposes, let us obtain a confidence interval for the long-run propensity in this model.

The estimated LRP in (10.19) is .073 − .0058 + .034 ≈ .101. However, we do not have enough information in (10.19) to obtain the standard error of this estimate. To obtain the standard error of the estimated LRP, we use the trick suggested in Section 4.4. Let $\theta_0 = \delta_0 + \delta_1 + \delta_2$ denote the LRP and write δ_0 in terms of θ_0, δ_1, and δ_2 as $\delta_0 = \theta_0 - \delta_1 - \delta_2$. Next, substitute for δ_0 in the model

$$gfr_t = \alpha_0 + \delta_0 pe_t + \delta_1 pe_{t-1} + \delta_2 pe_{t-2} + \ldots$$

to get

$$\begin{aligned} gfr_t &= \alpha_0 + (\theta_0 - \delta_1 - \delta_2)pe_t + \delta_1 pe_{t-1} + \delta_2 pe_{t-2} + \ldots \\ &= \alpha_0 + \theta_0 pe_t + \delta_1(pe_{t-1} - pe_t) + \delta_2(pe_{t-2} - pe_t) + \ldots. \end{aligned}$$

From this last equation, we can obtain $\hat{\theta}_0$ and its standard error by regressing gfr_t on pe_t, $(pe_{t-1} - pe_t)$, $(pe_{t-2} - pe_t)$, $ww2_t$, and $pill_t$. The coefficient and associated standard error on pe_t are what we need. Running this regression gives $\hat{\theta}_0 = .101$ as the coefficient on pe_t (as we already knew from above) and $\text{se}(\hat{\theta}_0) = .030$ [which we could not compute from (10.19)]. Therefore, the *t* statistic for $\hat{\theta}_0$ is about 3.37, so $\hat{\theta}_0$ is statistically different from zero at small significance levels. Even though none of the $\hat{\delta}_j$ is individually significant, the LRP is very significant. The 95% confidence interval for the LRP is about .041 to .160.

Whittington, Alm, and Peters (1990) allow for further lags but restrict the coefficients to help alleviate the multicollinearity problem that hinders estimation of the individual δ_j. (See Problem 10.6 for an example of how to do this.) For estimating the LRP, which would seem to be of primary interest here, such restrictions are unnecessary. Whittington, Alm, and Peters also control for additional variables, such as average female wage and the unemployment rate.

Binary explanatory variables are the key component in what is called an **event study**. In an event study, the goal is to see whether a particular event influences some outcome. Economists who study industrial organization have looked at the effects of certain events on firm stock prices. For example, Rose (1985) studied the effects of new trucking regulations on the stock prices of trucking companies.

A simple version of an equation used for such event studies is

$$R_t^f = \beta_0 + \beta_1 R_t^m + \beta_2 d_t + u_t,$$

where R_t^f is the stock return for firm f during period t (usually a week or a month), R_t^m is the market return (usually computed for a broad stock market index), and d_t is a dummy variable indicating when the event occurred. For example, if the firm is an airline, d_t might denote whether the airline experienced a publicized accident or near accident during week t. Including R_t^m in the equation controls for the possibility that broad market movements might coincide with airline accidents. Sometimes, multiple dummy variables are used. For example, if the event is the imposition of a new regulation that might affect a certain firm, we might include a dummy variable that is one for a few weeks before the regulation was publicly announced and a second dummy variable for a few weeks after the regulation was announced. The first dummy variable might detect the presence of inside information.

Before we give an example of an event study, we need to discuss the notion of an **index number** and the difference between nominal and real economic variables. An index number typically aggregates a vast amount of information into a single quantity. Index numbers are used regularly in time series analysis, especially in macroeconomic applications. An example of an index number is the index of industrial production (IIP), computed monthly by the Board of Governors of the Federal Reserve. The IIP is a measure of production across a broad range of industries, and, as such, its magnitude in a particular year has no quantitative meaning. In order to interpret the magnitude of the IIP, we must know the **base period** and the **base value**. In the 1997 *Economic Report of the President* (*ERP*), the base year is 1987, and the base value is 100. (Setting IIP to 100 in the base period is just a convention; it makes just as much sense to set IIP = 1 in 1987, and some indexes are defined with one as the base value.) Because the IIP was 107.7 in 1992, we can say that industrial production was 7.7% higher in 1992 than in 1987. We can use the IIP in any two years to compute the percentage difference in industrial output during those two years. For example, since IIP = 61.4 in 1970 and IIP = 85.7 in 1979, industrial production grew by about 39.6% during the 1970s.

It is easy to change the base period for any index number, and sometimes we must do this to give index numbers reported with different base years a common base year. For example, if we want to change the base year of the IIP from 1987 to 1982, we simply divide the IIP for each year by the 1982 value and then multiply by 100 to make the base period value 100. Generally, the formula is

$$newindex_t = 100(oldindex_t/oldindex_{\text{newbase}}), \tag{10.20}$$

where $oldindex_{\text{newbase}}$ is the original value of the index in the new base year. For example, with base year 1987, the IIP in 1992 is 107.7; if we change the base year to 1982, the IIP in 1992 becomes 100(107.7/81.9) = 131.5 (because the IIP in 1982 was 81.9).

Another important example of an index number is a *price index*, such as the consumer price index (CPI). We already used the CPI to compute annual inflation rates in Example 10.1. As with the industrial production index, the CPI is only meaningful when we compare it across different years (or months, if we are using monthly data). In the 1997 *ERP*, CPI = 38.8 in 1970, and CPI = 130.7 in 1990. Thus, the general price level grew by almost 237% over this twenty-year period. (In 1997, the CPI is defined so that its average in 1982, 1983, and 1984 equals 100; thus, the base period is listed as 1982–1984.)

In addition to being used to compute inflation rates, price indexes are necessary for turning a time series measured in *nominal dollars* (or *current dollars*) into *real dollars* (or *constant dollars*). Most economic behavior is assumed to be influenced by real, not nominal, variables. For example, classical labor economics assumes that labor supply is based on the real hourly wage, not the nominal wage. Obtaining the real wage from the nominal wage is easy if we have a price index such as the CPI. We must be a little careful to first divide the CPI by 100, so that the value in the base year is one. Then, if w denotes the average hourly wage in nominal dollars and p = CPI/100, the *real wage* is simply w/p. This wage is measured in dollars for the base period of the CPI. For example, in Table B-45 in the 1997 *ERP*, average hourly earnings are reported in nominal terms and in 1982 dollars (which means that the CPI used in computing the real wage had the base year 1982). This table reports that the nominal hourly wage in 1960 was \$2.09, but measured in 1982 dollars, the wage was \$6.79. The real hourly wage had peaked in 1973, at \$8.55 in 1982 dollars, and had fallen to \$7.40 by 1995. Thus, there has been a nontrivial decline in real wages over the past 20 years. (If we compare nominal wages from 1973 and 1995, we get a very misleading picture: \$3.94 in 1973 and \$11.44 in 1995. Since the real wage has actually fallen, the increase in the nominal wage is due entirely to inflation.)

Standard measures of economic output are in real terms. The most important of these is *gross domestic product*, or *GDP*. When growth in GDP is reported in the popular press, it is always *real* GDP growth. In the 1997 *ERP*, Table B-9, GDP is reported in billions of 1992 dollars. We used a similar measure of output, real gross national product, in Example 10.3.

Interesting things happen when real dollar variables are used in combination with natural logarithms. Suppose, for example, that average weekly hours worked are related to the real wage as

$$\log(hours) = \beta_0 + \beta_1 \log(w/p) + u.$$

Using the fact that $\log(w/p) = \log(w) - \log(p)$, we can write this as

$$\log(hours) = \beta_0 + \beta_1 \log(w) + \beta_2 \log(p) + u, \tag{10.21}$$

but with the restriction that $\beta_2 = -\beta_1$. Therefore, the assumption that only the real wage influences labor supply imposes a restriction on the parameters of model (10.21).

If $\beta_2 \neq -\beta_1$, then the price level has an effect on labor supply, something that can happen if workers do not fully understand the distinction between real and nominal wages.

There are many practical aspects to the actual computation of index numbers, but it would take us too far afield to cover those here. Detailed discussions of price indexes can be found in most intermediate macroeconomic texts, such as Mankiw (1994, Chapter 2). For us, it is important to be able to use index numbers in regression analysis. As mentioned earlier, since the magnitudes of index numbers are not especially informative, they often appear in logarithmic form, so that regression coefficients have percentage change interpretations.

We now give an example of an event study that also uses index numbers.

EXAMPLE 10.5
(Antidumping Filings and Chemical Imports)

Krupp and Pollard (1996) analyzed the effects of antidumping filings by U.S. chemical industries on imports of various chemicals. We focus here on one industrial chemical, barium chloride, a cleaning agent used in various chemical processes and in gasoline production. In the early 1980s, U.S. barium chloride producers believed that China was offering its U.S. imports at an unfairly low price (an action known as *dumping*), and the barium chloride industry filed a complaint with the U.S. International Trade Commission (ITC) in October 1983. The ITC ruled in favor of the U.S. barium chloride industry in October 1984. There are several questions of interest in this case, but we will touch on only a few of them. First, are imports unusually high in the period immediately preceding the initial filing? Second, do imports change noticeably after an antidumping filing? Finally, what is the reduction in imports after a decision in favor of the U.S. industry?

To answer these questions, we follow Krupp and Pollard by defining three dummy variables: *befile6* is equal to one during the six months before filing, *affile6* indicates the six months after filing, and *afdec6* denotes the six months after the positive decision. The dependent variable is the volume of imports of barium chloride from China, *chnimp*, which we use in logarithmic form. We include as explanatory variables, all in logarithmic form, an index of chemical production, *chempi* (to control for overall demand for barium chloride), the volume of gasoline production, *gas* (another demand variable), and an exchange rate index, *rtwex*, which measures the strength of the dollar against several other currencies. The chemical production index was defined to be 100 in June 1977. The analysis here differs somewhat from Krupp and Pollard in that we use natural logarithms of all variables (except the dummy variables, of course), and we include all three dummy variables in the same regression.

Using monthly data from February 1978 through December 1988 gives the following:

$$
\begin{aligned}
\log(\widehat{chnimp}) = &\underset{(21.05)}{-17.80} + \underset{(0.48)}{3.12} \log(chempi) + \underset{(.907)}{.196} \log(gas) \\
&+ \underset{(.400)}{.983} \log(rtwex) + \underset{(.261)}{.060}\, befile6 - \underset{(.264)}{.032}\, affile6 - \underset{(.286)}{.566}\, afdec6 \\
&n = 131,\ R^2 = .305,\ \bar{R}^2 = .271.
\end{aligned}
\tag{10.22}
$$

The equation shows that *befile6* is statistically insignificant, so there is no evidence that Chinese imports were unusually high during the six months before the suit was filed. Further, although the estimate on *affile6* is negative, the coefficient is small (indicating about a 3.2% fall in Chinese imports), and it is statistically very insignificant. The coefficient on *afdec6* shows a substantial fall in Chinese imports of barium chloride after the decision in favor of the U.S. industry, which is not surprising. Since the effect is so large, we compute the exact percentage change: $100[\exp(-.566) - 1] \approx -43.2\%$. The coefficient is statistically significant at the 5% level against a two-sided alternative.

The coefficient signs on the control variables are what we expect: an increase in overall chemical production increases the demand for the cleaning agent. Gasoline production does not affect Chinese imports significantly. The coefficient on log(*rtwex*) shows that an increase in the value of the dollar relative to other currencies increases the demand for Chinese imports, as is predicted by economic theory. (In fact, the elasticity is not statistically different from one. Why?)

Interactions among qualitative and quantitative variables are also used in time series analysis. An example with practical importance follows.

EXAMPLE 10.6
(Election Outcomes and Economic Performance)

Fair (1996) summarizes his work on explaining presidential election outcomes in terms of economic performance. He explains the proportion of the two-party vote going to the Democratic candidate using data for the years 1916 through 1992 (every four years) for a total of 20 observations. We estimate a simplified version of Fair's model (using variable names that are more descriptive than his):

$$\begin{aligned} demvote = \beta_0 + \beta_1 partyWH + \beta_2 incum + \beta_3 partyWH \cdot gnews \\ + \beta_4 partyWH \cdot inf + u, \end{aligned}$$

where *demvote* is the proportion of the two-party vote going to the Democratic candidate. The explanatory variable *partyWH* is similar to a dummy variable, but it takes on the value one if a Democrat is in the White House and -1 if a Republican is in the White House. Fair uses this variable to impose the restriction that the effect of a Republican being in the White House has the same magnitude but opposite sign as a Democrat being in the White House. This is a natural restriction since the party shares must sum to one, by definition. It also saves two degrees of freedom, which is important with so few observations. Similarly, the variable *incum* is defined to be one if a Democratic incumbent is running, -1 if a Republican incumbent is running, and zero otherwise. The variable *gnews* is the number of quarters during the current administration's first 15 (out of 16 total), where the quarterly growth in real per capita output was above 2.9% (at an annual rate), and *inf* is the average annual inflation rate over the first 15 quarters of the administration. See Fair (1996) for precise definitions.

Economists are most interested in the interaction terms *partyWH·gnews* and *partyWH·inf*. Since *partyWH* equals one when a Democrat is in the White House, β_3 measures the effect of good economic news on the party in power; we expect $\beta_3 > 0$. Similarly,

β_4 measures the effect that inflation has on the party in power. Because inflation during an administration is considered to be bad news, we expect $\beta_4 < 0$.

The estimated equation using the data in FAIR.RAW is

$$
\begin{aligned}
\widehat{demvote} = &\ \underset{(.012)}{.481} - \underset{(.0405)}{.0435}\ partyWH + \underset{(.0234)}{.0544}\ incum \\
&+ \underset{(.0041)}{.0108}\ partyWH \cdot gnews - \underset{(.0033)}{.0077}\ partyWH \cdot inf \\
&n = 20,\ R^2 = .663,\ \bar{R}^2 = .573.
\end{aligned}
\tag{10.23}
$$

All coefficients, except that on *partyWH*, are statistically significant at the 5% level. Incumbency is worth about 5.4 percentage points in the share of the vote. (Remember, *demvote* is measured as a proportion.) Further, the economic news variable has a positive effect: one more quarter of good news is worth about 1.1 percentage points. Inflation, as expected, has a negative effect: if average annual inflation is, say, two percentage points higher, the party in power loses about 1.5 percentage points of the two-party vote.

We could have used this equation to predict the outcome of the 1996 presidential election between Bill Clinton, the Democrat, and Bob Dole, the Republican. (The independent candidate, Ross Perot, is excluded because Fair's equation is for the two-party vote only.) Since Clinton ran as an incumbent, $partyWH = 1$ and $incum = 1$. To predict the election outcome, we need the variables *gnews* and *inf*. During Clinton's first 15 quarters in office, per capita real GDP exceeded 2.9% three times, so $gnews = 3$. Further, using the GDP price deflator reported in Table B-4 in the 1997 *ERP*, the average annual inflation rate (computed using Fair's formula) from the fourth quarter in 1991 to the third quarter in 1996 was 3.019. Plugging these into (10.23) gives

$$\widehat{demvote} = .481 - .0435 + .0544 + .0108(3) - .0077(3.019) \approx .5011.$$

Therefore, based on information known before the election in November, Clinton was predicted to receive a very slight majority of the two-party vote: about 50.1%. In fact, Clinton won more handily: his share of the two-party vote was 54.65%.

10.5 TRENDS AND SEASONALITY

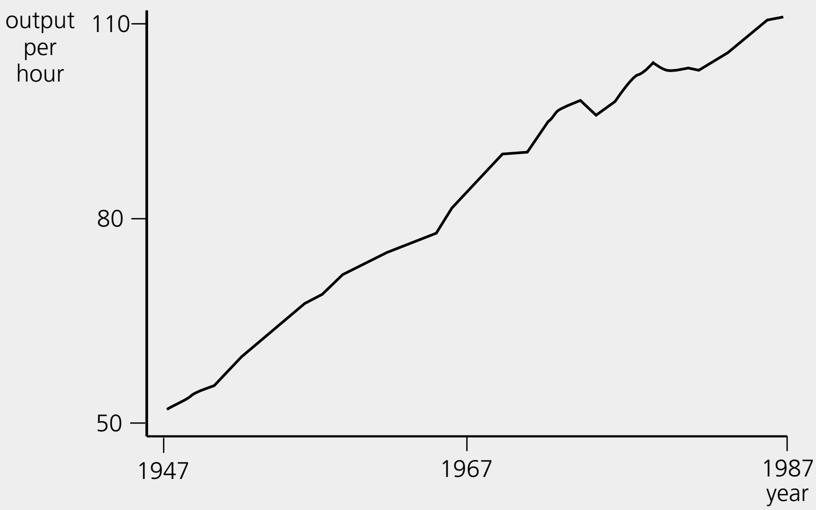
Characterizing Trending Time Series

Many economic time series have a common tendency of growing over time. We must recognize that some series contain a **time trend** in order to draw causal inference using time series data. Ignoring the fact that two sequences are trending in the same or opposite directions can lead us to falsely conclude that changes in one variable are actually caused by changes in another variable. In many cases, two time series processes appear to be correlated only because they are both trending over time for reasons related to other unobserved factors.

Figure 10.2 contains a plot of labor productivity (output per hour of work) in the United States for the years 1947 through 1987. This series displays a clear upward trend, which reflects the fact that workers have become more productive over time.

Figure 10.2

Output per labor hour in the United States during the years 1947–1987; 1977 = 100.



Other series, at least over certain time periods, have clear downward trends. Because positive trends are more common, we will focus on those during our discussion.

What kind of statistical models adequately capture trending behavior? One popular formulation is to write the series $\{y_t\}$ as

$$y_t = \alpha_0 + \alpha_1 t + e_t, \; t = 1,2, \ldots, \tag{10.24}$$

where, in the simplest case, $\{e_t\}$ is an independent, identically distributed (i.i.d.) sequence with $\mathrm{E}(e_t) = 0$, $\mathrm{Var}(e_t) = \sigma_e^2$. Note how the parameter α_1 multiplies time, t, resulting in a **linear time trend**. Interpreting α_1 in (10.24) is simple: holding all other factors (those in e_t) fixed, α_1 measures the change in y_t from one period to the next due to the passage of time: when $\Delta e_t = 0$,

$$\Delta y_t = y_t - y_{t-1} = \alpha_1.$$

Another way to think about a sequence that has a linear time trend is that its average value is a linear function of time:

$$\mathrm{E}(y_t) = \alpha_0 + \alpha_1 t. \tag{10.25}$$

If $\alpha_1 > 0$, then, on average, y_t is growing over time and therefore has an upward trend. If $\alpha_1 < 0$, then y_t has a downward trend. The values of y_t do not fall exactly on the line

QUESTION 10.4

In Example 10.4, we used the general fertility rate as the dependent variable in a finite distributed lag model. From 1950 through the mid-1980s, the *gfr* has a clear downward trend. Can a linear trend with $\alpha_1 < 0$ be realistic for all future time periods? Explain.

in (10.25) due to randomness, but the expected values are on the line. Unlike the mean, the variance of y_t is constant across time: $\text{Var}(y_t) = \text{Var}(e_t) = \sigma_e^2$.

If $\{e_t\}$ is an i.i.d. sequence, then $\{y_t\}$ is an independent, though not identically, distributed sequence. A more realistic characterization of trending time series allows $\{e_t\}$ to be correlated over time, but this does not change the flavor of a linear time trend. In fact, what is important for regression analysis under the classical linear model assumptions is that $\text{E}(y_t)$ is linear in t. When we cover large sample properties of OLS in Chapter 11, we will have to discuss how much temporal correlation in $\{e_t\}$ is allowed.

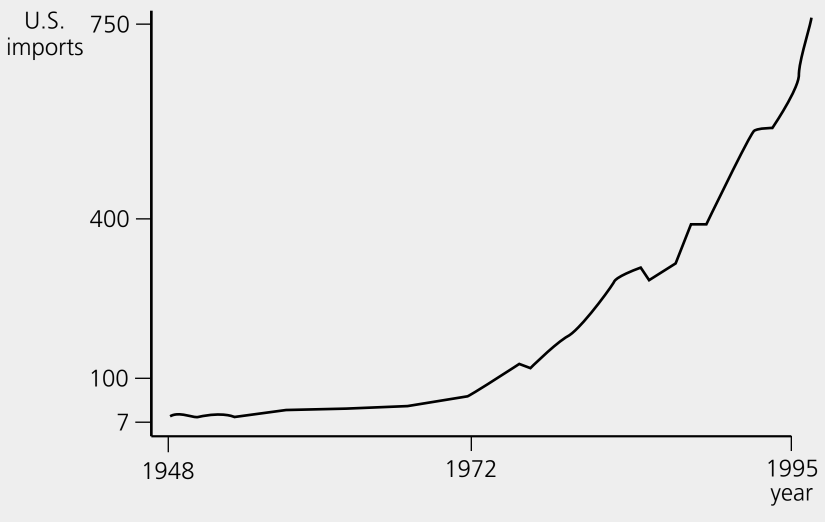
Many economic time series are better approximated by an **exponential trend**, which follows when a series has the same average growth rate from period to period. Figure 10.3 plots data on annual nominal imports for the United States during the years 1948 through 1995 (*ERP* 1997, Table B–101).

In the early years, we see that the change in the imports over each year is relatively small, whereas the change increases as time passes. This is consistent with a *constant average growth rate*: the percentage change is roughly the same in each period.

In practice, an exponential trend in a time series is captured by modeling the natural logarithm of the series as a linear trend (assuming that $y_t > 0$):

Figure 10.3

Nominal U.S. imports during the years 1948–1995 (in billions of U.S. dollars).



$$\log(y_t) = \beta_0 + \beta_1 t + e_t,\ t = 1,2, \ldots. \tag{10.26}$$

Exponentiating shows that y_t itself has an exponential trend: $y_t = \exp(\beta_0 + \beta_1 t + e_t)$. Because we will want to use exponentially trending time series in linear regression models, (10.26) turns out to be the most convenient way for representing such series.

How do we interpret β_1 in (10.26)? Remember that, for small changes, $\Delta\log(y_t) = \log(y_t) - \log(y_{t-1})$ is approximately the proportionate change in y_t:

$$\Delta\log(y_t) \approx (y_t - y_{t-1})/y_{t-1}. \tag{10.27}$$

The right-hand side of (10.27) is also called the **growth rate** in y from period $t - 1$ to period t. To turn the growth rate into a percent, we simply multiply by 100. If y_t follows (10.26), then, taking changes and setting $\Delta e_t = 0$,

$$\Delta\log(y_t) = \beta_1, \text{ for all } t. \tag{10.28}$$

In other words, β_1 is approximately the average per period growth rate in y_t. For example, if t denotes year and $\beta_1 = .027$, then y_t grows about 2.7% per year on average.

Although linear and exponential trends are the most common, time trends can be more complicated. For example, instead of the linear trend model in (10.24), we might have a quadratic time trend:

$$y_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + e_t. \tag{10.29}$$

If α_1 and α_2 are positive, then the slope of the trend is increasing, as is easily seen by computing the approximate slope (holding e_t fixed):

$$\frac{\Delta y_t}{\Delta t} \approx \alpha_1 + 2\alpha_2 t. \tag{10.30}$$

[If you are familiar with calculus, you recognize the right-hand side of (10.30) as the derivative of $\alpha_0 + \alpha_1 t + \alpha_2 t^2$ with respect to t.] If $\alpha_1 > 0$, but $\alpha_2 < 0$, the trend has a hump shape. This may not be a very good description of certain trending series because it requires an increasing trend to be followed, eventually, by a decreasing trend. Nevertheless, over a given time span, it can be a flexible way of modeling time series that have more complicated trends than either (10.24) or (10.26).

Using Trending Variables in Regression Analysis

Accounting for explained or explanatory variables that are trending is fairly straightforward in regression analysis. First, nothing about trending variables necessarily violates the classical linear model assumptions, TS.1 through TS.6. However, we must be careful to allow for the fact that unobserved, trending factors that affect y_t might also be correlated with the explanatory variables. If we ignore this possibility, we may find a spurious relationship between y_t and one or more explanatory variables. The phenomenon of finding a relationship between two or more trending variables simply because each is growing over time is an example of **spurious regression**. Fortunately, adding a time trend eliminates this problem.

For concreteness, consider a model where two observed factors, x_{t1} and x_{t2}, affect y_t. In addition, there are unobserved factors that are systematically growing or shrinking over time. A model that captures this is

$$y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \beta_3 t + u_t. \tag{10.31}$$

This fits into the multiple linear regression framework with $x_{t3} = t$. Allowing for the trend in this equation explicitly recognizes that y_t may be growing ($\beta_3 > 0$) or shrinking ($\beta_3 < 0$) over time for reasons essentially unrelated to x_{t1} and x_{t2}. If (10.31) satisfies assumptions TS.1, TS.2, and TS.3, then omitting t from the regression and regressing y_t on x_{t1}, x_{t2} will generally yield biased estimators of β_1 and β_2: we have effectively omitted an important variable, t, from the regression. This is especially true if x_{t1} and x_{t2} are themselves trending, because they can then be highly correlated with t. The next example shows how omitting a time trend can result in spurious regression.

EXAMPLE 10.7
(Housing Investment and Prices)

The data in HSEINV.RAW are annual observations on housing investment and a housing price index in the United States for 1947 through 1988. Let *invpc* denote real per capita housing investment (in thousands of dollars) and let *price* denote a housing price index (equal to one in 1982). A simple regression in constant elasticity form, which can be thought of as a supply equation for housing stock, gives

$$\widehat{\log(invpc)} = \underset{(.043)}{-.550} + \underset{(0.382)}{1.241} \log(price)$$
$$n = 42, R^2 = .208, \bar{R}^2 = .189. \tag{10.32}$$

The elasticity of per capita investment with respect to price is very large and statistically significant; it is not statistically different from one. We must be careful here. Both *invpc* and *price* have upward trends. In particular, if we regress log(*invpc*) on t, we obtain a coefficient on the trend equal to .0081 (standard error = .0018); the regression of log(*price*) on t yields a trend coefficient equal to .0044 (standard error = .0004). While the standard errors on the trend coefficients are not necessarily reliable—these regressions tend to contain substantial serial correlation—the coefficient estimates do reveal upward trends.

To account for the trending behavior of the variables, we add a time trend:

$$\widehat{\log(invpc)} = \underset{(.136)}{-.913} - \underset{(.679)}{.381} \log(price) + \underset{(.0035)}{.0098}\, t$$
$$n = 42, R^2 = .341, \bar{R}^2 = .307. \tag{10.33}$$

The story is much different now: the estimated price elasticity is negative and not statistically different from zero. The time trend is statistically significant, and its coefficient implies an approximate 1% increase in *invpc* per year, on average. From this analysis, we cannot conclude that real per capita housing investment is influenced at all by price. There are other factors, captured in the time trend, that affect *invpc*, but we have not modeled these. The results in (10.32) show a spurious relationship between *invpc* and *price* due to the fact that price is also trending upward over time.

In some cases, adding a time trend can make a key explanatory variable *more* significant. This can happen if the dependent and independent variables have different kinds of trends (say, one upward and one downward), but movement in the independent variable *about* its trend line causes movement in the dependent variable away from its trend line.

EXAMPLE 10.8
(Fertility Equation)

If we add a linear time trend to the fertility equation (10.18), we obtain

$$\begin{aligned} \widehat{gfr}_t = &\ 111.77 + .279\ pe_t - 35.59\ ww2_t + .997\ pill_t - 1.15\ t \\ &\ (3.36)\quad (.040) \qquad (6.30) \qquad (6.626) \qquad (0.19) \\ &\ n = 72, R^2 = .662, \bar{R}^2 = .642. \end{aligned} \tag{10.34}$$

The coefficient on *pe* is more than triple the estimate from (10.18), and it is much more statistically significant. Interestingly, *pill* is not significant once an allowance is made for a linear trend. As can be seen by the estimate, *gfr* was falling, on average, over this period, other factors being equal.

Since the general fertility rate exhibited both upward and downward trends during the period from 1913 through 1984, we can see how robust the estimated effect of *pe* is when we use a quadratic trend:

$$\begin{aligned} \widehat{gfr}_t = &\ 124.09 + .348\ pe_t - 35.88\ ww2_t - 10.12\ pill_t \\ &\ (4.36)\quad (.040) \qquad (5.71) \qquad (6.34) \\ &\ - 2.53\ t + .0196\ t^2 \\ &\ \quad (0.39) \quad (.0050) \\ &\ n = 72, R^2 = .727, \bar{R}^2 = .706. \end{aligned} \tag{10.35}$$

The coefficient on *pe* is even larger and more statistically significant. Now, *pill* has the expected negative effect and is marginally significant, and both trend terms are statistically significant. The quadratic trend is a flexible way to account for the unusual trending behavior of *gfr*.

You might be wondering in Example 10.8: Why stop at a quadratic trend? Nothing prevents us from adding, say, t^3 as an independent variable, and, in fact, this might be warranted (see Exercise 10.12). But we have to be careful not to get carried away when including trend terms in a model. We want relatively simple trends that capture broad movements in the dependent variable that are not explained by the independent variables in the model. If we include enough polynomial terms in t, then we can track any series pretty well. But this offers little help in finding which explanatory variables affect y_t.

A Detrending Interpretation of Regressions with a Time Trend

Including a time trend in a regression model creates a nice interpretation in terms of **detrending** the original data series before using them in regression analysis. For concreteness, we focus on model (10.31), but our conclusions are much more general.

When we regress y_t on x_{t1}, x_{t2} and t, we obtain the fitted equation

$$\hat{y}_t = \hat{\beta}_0 + \hat{\beta}_1 x_{t1} + \hat{\beta}_2 x_{t2} + \hat{\beta}_3 t. \quad \textbf{(10.36)}$$

We can extend the results on the partialling out interpretation of OLS that we covered in Chapter 3 to show that $\hat{\beta}_1$ and $\hat{\beta}_2$ can be obtained as follows.

(i) Regress each of y_t, x_{t1} and x_{t2} on a constant and the time trend t and save the residuals, say $\ddot{y}_t$, $\ddot{x}_{t1}$, $\ddot{x}_{t2}$, $t = 1,2, \ldots, n$. For example,

$$\ddot{y}_t = y_t - \hat{\alpha}_0 - \hat{\alpha}_1 t.$$

Thus, we can think of $\ddot{y}_t$ as being *linearly detrended.* In detrending y_t, we have estimated the model

$$y_t = \alpha_0 + \alpha_1 t + e_t$$

by OLS; the residuals from this regression, $\hat{e}_t = \ddot{y}_t$, have the time trend removed (at least in the sample). A similar interpretation holds for $\ddot{x}_{t1}$ and $\ddot{x}_{t2}$.

(ii) Run the regression of

$$\ddot{y}_t \text{ on } \ddot{x}_{t1}, \ddot{x}_{t2}. \quad \textbf{(10.37)}$$

(No intercept is necessary, but including an intercept affects nothing: the intercept will be estimated to be zero.) This regression exactly yields $\hat{\beta}_1$ and $\hat{\beta}_2$ from (10.36).

This means that the estimates of primary interest, $\hat{\beta}_1$ and $\hat{\beta}_2$, can be interpreted as coming from a regression *without* a time trend, but where we first detrend the dependent variable and all other independent variables. The same conclusion holds with any number of independent variables and if the trend is quadratic or of some other polynomial degree.

If t is omitted from (10.36), then no detrending occurs, and y_t might seem to be related to one or more of the x_{tj} simply because each contains a trend; we saw this in Example 10.7. If the trend term is statistically significant, and the results change in important ways when a time trend is added to a regression, then the initial results without a trend should be treated with suspicion.

The interpretation of $\hat{\beta}_1$ and $\hat{\beta}_2$ shows that it is a good idea to include a trend in the regression if any independent variable is trending, even if y_t is not. If y_t has no noticeable trend, but, say, x_{t1} is growing over time, then excluding a trend from the regression may make it look as if x_{t1} has no effect on y_t, even though movements of x_{t1} about its trend may affect y_t. This will be captured if t is included in the regression.

EXAMPLE 10.9
(Puerto Rican Employment)

When we add a linear trend to equation (10.17), the estimates are

$$
\begin{aligned}
\log(\widehat{prepop}_t) = {} & \underset{(1.30)}{-8.70} - \underset{(.044)}{.169} \log(mincov_t) + \underset{(0.18)}{1.06} \log(usgnp_t) \\
& - \underset{(.005)}{.032}\, t \\
& n = 38,\ R^2 = .847,\ \bar{R}^2 = .834.
\end{aligned}
\tag{10.38}
$$

The coefficient on log(*usgnp*) has changed dramatically: from −.012 and insignificant to 1.06 and very significant. The coefficient on the minimum wage has changed only slightly, although the standard error is notably smaller, making log(*mincov*) more significant than before.

The variable $prepop_t$ displays no clear upward or downward trend, but log(*usgnp*) has an upward, linear trend. (A regression of log(*usgnp*) on t gives an estimate of about .03, so that *usgnp* is growing by about 3% per year over the period.) We can think of the estimate 1.06 as follows: when *usgnp* increases by 1% *above* its long-run trend, *prepop* increases by about 1.06%.

Computing *R*-squared when the Dependent Variable is Trending

R-squareds in time series regressions are often very high, especially compared with typical *R*-squareds for cross-sectional data. Does this mean that we learn more about factors affecting y from time series data? Not necessarily. On one hand, time series data often come in aggregate form (such as average hourly wages in the U.S. economy), and aggregates are often easier to explain than outcomes on individuals, families, or firms, which is often the nature of cross-sectional data. But the usual and adjusted *R*-squares for time series regressions can be artificially high when the dependent variable is trending. Remember that R^2 is a measure of how large the error variance is relative to the variance of y. The formula for the adjusted *R*-squared shows this directly:

$$\bar{R}^2 = 1 - (\hat{\sigma}_u^2/\hat{\sigma}_y^2),$$

where $\hat{\sigma}_u^2$ is the unbiased estimator of the error variance, $\hat{\sigma}_y^2 = \text{SST}/(n-1)$, and $\text{SST} = \sum_{t=1}^{n} (y_t - \bar{y})^2$. Now, estimating the error variance when y_t is trending is no problem, provided a time trend is included in the regression. However, when $\text{E}(y_t)$ follows, say, a linear time trend [see (10.24)], $\text{SST}/(n-1)$ is no longer an unbiased or consistent estimator of $\text{Var}(y_t)$. In fact, $\text{SST}/(n-1)$ can substantially overestimate the variance in y_t, because it does not account for the trend in y_t.

When the dependent variable satisfies linear, quadratic, or any other polynomial trends, it is easy to compute a goodness-of-fit measure that first nets out the effect of any time trend on y_t. The simplest method is to compute the usual R-squared in a regression where the dependent variable has already been detrended. For example, if the model is (10.31), then we first regress y_t on t and obtain the residuals $\ddot{y}_t$. Then, we regress

$$\ddot{y}_t \text{ on } x_{t1}, x_{t2}, \text{ and } t. \tag{10.39}$$

The R-squared from this regression is

$$1 - \frac{\text{SSR}}{\sum_{t=1}^{n} \ddot{y}_t^2}, \tag{10.40}$$

where SSR is identical to the sum of squared residuals from (10.36). Since $\sum_{t=1}^{n} \ddot{y}_t^2 \leq \sum_{t=1}^{n} (y_t - \bar{y})^2$ (and usually the inequality is strict), the R-squared from (10.40) is no greater than, and usually less than, the R-squared from (10.36). (The sum of squared residuals is identical in both regressions.) When y_t contains a strong linear time trend, (10.40) can be much less than the usual R-squared.

The R-squared in (10.40) better reflects how well x_{t1} and x_{t2} explain y_t, because it nets out the effect of the time trend. After all, we can always explain a trending variable with some sort of trend, but this does not mean we have uncovered any factors that cause movements in y_t. An adjusted R-squared can also be computed based on (10.40): divide SSR by $(n - 4)$ because this is the *df* in (10.36) and divide $\sum_{t=1}^{n} \ddot{y}_t^2$ by $(n - 2)$, as there are two trend parameters estimated in detrending y_t. In general, SSR is divided by the *df* in the usual regression (that includes any time trends), and $\sum_{t=1}^{n} \ddot{y}_t^2$ is divided by $(n - p)$, where p is the number of trend parameters estimated in detrending y_t. See Wooldridge (1991a) for further discussion on computing goodness-of-fit measures with trending variables.

EXAMPLE 10.10
(Housing Investment)

In Example 10.7, we saw that including a linear time trend along with log(*price*) in the housing investment equation had a substantial effect on the price elasticity. But the R-squared from regression (10.33), taken literally, says that we are "explaining" 34.1% of the variation in log(*invpc*). This is misleading. If we first detrend log(*invpc*) and regress the detrended variable on log(*price*) and t, the R-squared becomes .008, and the adjusted R-squared is actually negative. Thus, movements in log(*price*) about its trend have virtually no explanatory power for movements in log(*invpc*) about its trend. This is consistent with the fact that the t statistic on log(*price*) in equation (10.33) is very small.

Before leaving this subsection, we must make a final point. In computing the R-squared form of an F statistic for testing multiple hypotheses, we just use the usual R-squareds without any detrending. Remember, the R-squared form of the F statistic is just a computational device, and so the usual formula is always appropriate.

Seasonality

If a time series is observed at monthly or quarterly intervals (or even weekly or daily), it may exhibit **seasonality**. For example, monthly housing starts in the Midwest are strongly influenced by weather. While weather patterns are somewhat random, we can be sure that the weather during January will usually be more inclement than in June, and so housing starts are generally higher in June than in January. One way to model this phenomenon is to allow the expected value of the series, y_t, to be different in each month. As another example, retail sales in the fourth quarter are typically higher than in the previous three quarters because of the Christmas holiday. Again, this can be captured by allowing the average retail sales to differ over the course of a year. This is in addition to possibly allowing for a trending mean. For example, retail sales in the most recent first quarter were higher than retail sales in the fourth quarter from 30 years ago, because retail sales have been steadily growing. Nevertheless, if we compare average sales within a typical year, the seasonal holiday factor tends to make sales larger in the fourth quarter.

Even though many monthly and quarterly data series display seasonal patterns, not all of them do. For example, there is no noticeable seasonal pattern in monthly interest or inflation rates. In addition, series that do display seasonal patterns are often **seasonally adjusted** before they are reported for public use. A seasonally adjusted series is one that, in principle, has had the seasonal factors removed from it. Seasonal adjustment can be done in a variety of ways, and a careful discussion is beyond the scope of this text. [See Harvey (1990) and Hylleberg (1986) for detailed treatments.]

Seasonal adjustment has become so common that it is not possible to get seasonally unadjusted data in many cases. Quarterly U.S. GDP is a leading example. In the annual *Economic Report of the President*, many macroeconomic data sets reported at monthly frequencies (at least for the most recent years) and those that display seasonal patterns are all seasonally adjusted. The major sources for macroeconomic time series, including *Citibase*, also seasonally adjust many of the series. Thus, the scope for using our own seasonal adjustment is often limited.

Sometimes, we do work with seasonally unadjusted data, and it is useful to know that simple methods are available for dealing with seasonality in regression models. Generally, we can include a set of **seasonal dummy variables** to account for seasonality in the dependent variable, the independent variables, or both.

The approach is simple. Suppose that we have monthly data, and we think that seasonal patterns within a year are roughly constant across time. For example, since Christmas always comes at the same time of year, we can expect retail sales to be, on average, higher in months late in the year than in earlier months. Or, since weather patterns are broadly similar across years, housing starts in the Midwest will be higher on average during the summer months than the winter months. A general model for monthly data that captures these phenomena is

$$y_t = \beta_0 + \delta_1 feb_t + \delta_2 mar_t + \delta_3 apr_t + \ldots + \delta_{11} dec_t + \beta_1 x_{t1} + \ldots + \beta_k x_{tk} + u_t, \quad \textbf{(10.41)}$$

where feb_t, mar_t, …, dec_t are dummy variables indicating whether time period t corresponds to the appropriate month. In this formulation, January is the base month, and β_0 is the intercept for January. If there is no seasonality in y_t, once the x_{tj} have been controlled for, then δ_1 through δ_{11} are all zero. This is easily tested via an F test.

QUESTION 10.5

In equation (10.41), what is the intercept for March? Explain why seasonal dummy variables satisfy the strict exogeneity assumption.

EXAMPLE 10.11
(Effects of Antidumping Filings)

In Example 10.5, we used monthly data that have not been seasonally adjusted. Therefore, we should add seasonal dummy variables to make sure none of the important conclusions changes. It could be that the months just before the suit was filed are months where imports are higher or lower, on average, than in other months. When we add the 11 monthly dummy variables as in (10.41) and test their joint significance, we obtain p-value = .59, and so the seasonal dummies are jointly insignificant. In addition, nothing important changes in the estimates once statistical significance is taken into account. Krupp and Pollard (1996) actually used three dummy variables for the seasons (fall, spring, and summer, with winter as the base season), rather than a full set of monthly dummies; the outcome is essentially the same.

If the data are quarterly, then we would include dummy variables for three of the four quarters, with the omitted category being the base quarter. Sometimes, it is useful to interact seasonal dummies with some of the x_{tj} to allow the effect of x_{tj} on y_t to differ across the year.

Just as including a time trend in a regression has the interpretation of initially detrending the data, including seasonal dummies in a regression can be interpreted as **deseasonalizing** the data. For concreteness, consider equation (10.41) with $k = 2$. The OLS slope coefficients $\hat{\beta}_1$ and $\hat{\beta}_2$ on x_1 and x_2 can be obtained as follows:

(i) Regress each of y_t, x_{t1} and x_{t2} on a constant and the monthly dummies, feb_t, mar_t, …, dec_t, and save the residuals, say $\ddot{y}_t$, $\ddot{x}_{t1}$ and $\ddot{x}_{t2}$, for all $t = 1,2, \ldots, n$. For example,

$$\ddot{y}_t = y_t - \hat{\alpha}_0 - \hat{\alpha}_1 feb_t - \hat{\alpha}_2 mar_t - \ldots - \hat{\alpha}_{11} dec_t.$$

This is one method of deseasonalizing a monthly time series. A similar interpretation holds for $\ddot{x}_{t1}$ and $\ddot{x}_{t2}$.

(ii) Run the regression, without the monthly dummies, of $\ddot{y}_t$ on $\ddot{x}_{t1}$ and $\ddot{x}_{t2}$ [just as in (10.37)]. This gives $\hat{\beta}_1$ and $\hat{\beta}_2$.

In some cases, if y_t has pronounced seasonality, a better goodness-of-fit measure is an R-squared based on the deseasonalized y_t. This nets out any seasonal effects that are not explained by the x_{tj}. Specific degrees of freedom ajustments are discussed in Wooldridge (1991a).

Time series exhibiting seasonal patterns can be trending as well, in which case, we should estimate a regression model with a time trend and seasonal dummy variables. The regressions can then be interpreted as regressions using both detrended and deseasonalized series. Goodness-of-fit statistics are discussed in Wooldridge (1991a): essentially, we detrend and deasonalize y_t by regressing on both a time trend and seasonal dummies before computing R-squared.

SUMMARY

In this chapter, we have covered basic regression analysis with time series data. Under assumptions that parallel those for cross-sectional analysis, OLS is unbiased (under TS.1 through TS.3), OLS is BLUE (under TS.1 through TS.5), and the usual OLS standard errors, t statistics, and F statistics can be used for statistical inference (under TS.1 through TS.6). Because of the temporal correlation in most time series data, we must explicitly make assumptions about how the errors are related to the explanatory variables in all time periods and about the temporal correlation in the errors themselves. The classical linear model assumptions can be pretty restrictive for time series applications, but they are a natural starting point. We have applied them to both static regression and finite distributed lag models.

Logarithms and dummy variables are used regularly in time series applications and in event studies. We also discussed index numbers and time series measured in terms of nominal and real dollars.

Trends and seasonality can be easily handled in a multiple regression framework by including time and seasonal dummy variables in our regression equations. We presented problems with the usual R-squared as a goodness-of-fit measure and suggested some simple alternatives based on detrending or deseasonalizing.

KEY TERMS

Autocorrelation
Base Period
Base Value
Contemporaneously Exogenous
Deseasonalizing
Detrending
Event Study
Exponential Trend
Finite Distributed Lag (FDL) Model
Growth Rate
Impact Multiplier
Impact Propensity
Index Number
Lag Distribution
Linear Time Trend
Long-Run Elasticity
Long-Run Multiplier
Long-Run Propensity (LRP)
Seasonal Dummy Variables
Seasonality
Seasonally Adjusted
Serial Correlation
Short-Run Elasticity
Spurious Regression
Static Model
Stochastic Process
Strictly Exogenous
Time Series Process
Time Trend

PROBLEMS

10.1 Decide if you agree or disagree with each of the following statements and give a brief explanation of your decision:

(i) Like cross-sectional observations, we can assume that most time series observations are independently distributed.

(ii) The OLS estimator in a time series regression is unbiased under the first three Gauss-Markov assumptions.

(iii) A trending variable cannot be used as the dependent variable in multiple regression analysis.

(iv) Seasonality is not an issue when using annual time series observations.

10.2 Let $gGDP_t$ denote the annual percentage change in gross domestic product and let int_t denote a short-term interest rate. Suppose that $gGDP_t$ is related to interest rates by

$$gGDP_t = \alpha_0 + \delta_0 int_t + \delta_1 int_{t-1} + u_t,$$

where u_t is uncorrelated with int_t, int_{t-1}, and all other past values of interest rates. Suppose that the Federal Reserve follows the policy rule:

$$int_t = \gamma_0 + \gamma_1(gGDP_{t-1} - 3) + v_t,$$

where $\gamma_1 > 0$. (When last year's GDP growth is above 3%, the Fed increases interest rates to prevent an "overheated" economy.) If v_t is uncorrelated with all past values of int_t and u_t, argue that int_t must be correlated with u_{t-1}. (*Hint*: Lag the first equation for one time period and substitute for $gGDP_{t-1}$ in the second equation.) Which Gauss-Markov assumption does this violate?

10.3 Suppose y_t follows a second order FDL model:

$$y_t = \alpha_0 + \delta_0 z_t + \delta_1 z_{t-1} + \delta_2 z_{t-2} + u_t.$$

Let z^* denote the *equilibrium value* of z_t and let y^* be the equilibrium value of y_t, such that

$$y^* = \alpha_0 + \delta_0 z^* + \delta_1 z^* + \delta_2 z^*.$$

Show that the change in y^*, due to a change in z^*, equals the long-run propensity times the change in z^*:

$$\Delta y^* = LRP \cdot \Delta z^*.$$

This gives an alternative way of interpreting the LRP.

10.4 When the three event indicators *befile6*, *affile6*, and *afdec6* are dropped from equation (10.22), we obtain $R^2 = .281$ and $\bar{R}^2 = .264$. Are the event indicators jointly significant at the 10% level?

10.5 Suppose you have quarterly data on new housing starts, interest rates, and real per capita income. Specify a model for housing starts that accounts for possible trends and seasonality in the variables.

10.6 In Example 10.4, we saw that our estimates of the individual lag coefficients in a distributed lag model were very imprecise. One way to alleviate the multicollinearity problem is to assume that the δ_j follow a relatively simple pattern. For concreteness, consider a model with four lags:

$$y_t = \alpha_0 + \delta_0 z_t + \delta_1 z_{t-1} + \delta_2 z_{t-2} + \delta_3 z_{t-3} + \delta_4 z_{t-4} + u_t.$$

Now, let us assume that the δ_j follow a quadratic in the lag, j:

$$\delta_j = \gamma_0 + \gamma_1 j + \gamma_2 j^2,$$

for parameters γ_0, γ_1, and γ_2. This is an example of a *polynomial distributed lag (PDL) model.*

(i) Plug the formula for each δ_j into the distributed lag model and write the model in terms of the parameters γ_h, for $h = 0,1,2$.

(ii) Explain the regression you would run to estimate the γ_h.

(iii) The polynomial distributed lag model is a restricted version of the general model. How many restrictions are imposed? How would you test these? (*Hint*: Think F test.)

COMPUTER EXERCISES

10.7 In October 1979, the Federal Reserve changed its policy of targeting the money supply and instead began to focus directly on short-term interest rates. Using the data in INTDEF.RAW, define a dummy variable equal to one for years after 1979. Include this dummy in equation (10.15) to see if there is a shift in the interest rate equation after 1979. What do you conclude?

10.8 Use the data in BARIUM.RAW for this exercise.

(i) Add a linear time trend to equation (10.22). Are any variables, other than the trend, statistically significant?

(ii) In the equation estimated in part (i), test for joint significance of all variables except the time trend. What do you conclude?

(iii) Add monthly dummy variables to this equation and test for seasonality. Does including the monthly dummies change any other estimates or their standard errors in important ways?

10.9 Add the variable log(*prgnp*) to the minimum wage equation in (10.38). Is this variable significant? Interpret the coefficient. How does adding log(*prgnp*) affect the estimated minimum wage effect?

10.10 Use the data in FERTIL3.RAW to verify that the standard error for the LRP in equation (10.19) is about .030.

10.11 Use the data in EZANDERS.RAW for this exercise. The data are on monthly unemployment claims in Anderson Township in Indiana, from January 1980 through November 1988. In 1984, an enterprise zone (EZ) was located in Anderson (as well as other cities in Indiana). [See Papke (1994) for details.]

(i) Regress log(*uclms*) on a linear time trend and 11 monthly dummy variables. What was the overall trend in unemployment claims over this period? (Interpret the coefficient on the time trend.) Is there evidence of seasonality in unemployment claims?

(ii) Add *ez*, a dummy variable equal to one in the months Anderson had an EZ, to the regression in part (i). Does having the enterprise zone seem to decrease unemployment claims? By how much? [You should use formula (7.10) from Chapter 7.]
(iii) What assumptions do you need to make to attribute the effect in part (ii) to the creation of an EZ?

10.12 Use the data in FERTIL3.RAW for this exercise.
(i) Regress gfr_t on t and t^2 and save the residuals. This gives a detrended gfr_t, say $\ddot{gfr}_t$.
(ii) Regress $\ddot{gfr}_t$ on all of the variables in equation (10.35), including t and t^2. Compare the R-squared with that from (10.35). What do you conclude?
(iii) Reestimate equation (10.35) but add t^3 to the equation. Is this additional term statistically significant?

10.13 Use the data set CONSUMP.RAW for this exercise.
(i) Estimate a simple regression model relating the growth in real per capita consumption (of nondurables and services) to the growth in real per capita disposable income. Use the change in the logarithms in both cases. Report the results in the usual form. Interpret the equation and discuss statistical significance.
(ii) Add a lag of the growth in real per capita disposable income to the equation from part (i). What do you conclude about adjustment lags in consumption growth?
(iii) Add the real interest rate to the equation in part (i). Does it affect consumption growth?

10.14 Use the data in FERTIL3.RAW for this exercise.
(i) Add pe_{t-3} and pe_{t-4} to equation (10.19). Test for joint significance of these lags.
(ii) Find the estimated long-run propensity and its standard error in the model from part (i). Compare these with those obtained from equation (10.19).
(iii) Estimate the polynomial distributed lag model from Problem 10.6. Find the estimated LRP and compare this with what is obtained from the unrestricted model.

10.15 Use the data in VOLAT.RAW for this exercise. The variable *rsp500* is the monthly return on the Standard & Poors 500 stock market index, at an annual rate. (This includes price changes as well as dividends.) The variable *i3* is the return on three-month T-bills, and *pcip* is the percentage change in industrial production; these are also at an annual rate.
(i) Consider the equation

$$rsp500_t = \beta_0 + \beta_1 pcip_t + \beta_2 i3_t + u_t.$$

What signs do you think β_1 and β_2 should have?
(ii) Estimate the previous equation by OLS, reporting the results in standard form. Interpret the signs and magnitudes of the coefficients.

(iii) Which of the variables is statistically significant?

(iv) Does your finding from part (iii) imply that the return on the S&P 500 is predictable? Explain.

10.16 Consider the model estimated in (10.15); use the data in INTDEF.RAW.

(i) Find the correlation between *inf* and *def* over this sample period and comment.

(ii) Add a single lag of *inf* and *def* to the equation and report the results in the usual form.

(iii) Compare the estimated LRP for the effect of inflation from that in equation (10.15). Are they vastly different?

(iv) Are the two lags in the model jointly significant at the 5% level?

Chapter **Eleven**

Further Issues in Using OLS with Time Series Data

In Chapter 10, we discussed the finite sample properties of OLS for time series data under increasingly stronger sets of assumptions. Under the full set of classical linear model assumptions for time series, TS.1 through TS.6, OLS has *exactly* the same desirable properties that we derived for cross-sectional data. Likewise, statistical inference is carried out in the same way as it was for cross-sectional analysis.

From our cross-sectional analysis in Chapter 5, we know that there are good reasons for studying the large sample properties of OLS. For example, if the error terms are not drawn from a normal distribution, then we must rely on the central limit theorem to justify the usual OLS test statistics and confidence intervals.

Large sample analysis is even more important in time series contexts. (This is somewhat ironic given that large time series samples can be difficult to come by; but we often have no choice other than to rely on large sample approximations.) In Section 10.3, we explained how the strict exogeneity assumption (TS.2) might be violated in static and distributed lag models. As we will show in Section 11.2, models with lagged dependent variables must violate Assumption TS.2.

Unfortunately, large sample analysis for time series problems is fraught with many more difficulties than it was for cross-sectional analysis. In Chapter 5, we obtained the large sample properties of OLS in the context of random sampling. Things are more complicated when we allow the observations to be correlated across time. Nevertheless, the major limit theorems hold for certain, although not all, time series processes. The key is whether the correlation between the variables at different time periods tends to zero quickly enough. Time series that have substantial temporal correlation require special attention in regression analysis. This chapter will alert you to certain issues pertaining to such series in regression analysis.

11.1 STATIONARY AND WEAKLY DEPENDENT TIME SERIES

In this section, we present the key concepts that are needed to apply the usual large sample approximations in regression analysis with time series data. The details are not as important as a general understanding of the issues.

Stationary and Nonstationary Time Series

Historically, the notion of a **stationary process** has played an important role in the analysis of time series. A stationary time series process is one whose probability distributions are stable over time in the following sense: if we take any collection of random variables in the sequence and then shift that sequence ahead h time periods, the joint probability distribution must remain unchanged. A formal definition of stationarity follows.

STATIONARY STOCHASTIC PROCESS: The stochastic process $\{x_t: t = 1,2, \dots\}$ is *stationary* if for every collection of time indices $1 \leq t_1 < t_2 < \dots < t_m$, the joint distribution of $(x_{t_1}, x_{t_2}, \dots, x_{t_m})$ is the same as the joint distribution of $(x_{t_1+h}, x_{t_2+h}, \dots, x_{t_m+h})$ for all integers $h \geq 1$.

This definition is a little abstract, but its meaning is pretty straightforward. One implication (by choosing $m = 1$ and $t_1 = 1$) is that x_t has the same distribution as x_1 for all $t = 2,3, \dots$. In other words, the sequence $\{x_t: t = 1,2, \dots\}$ is *identically distributed*. Stationarity requires even more. For example, the joint distribution of (x_1,x_2) (the first two terms in the sequence) must be the same as the joint distribution of (x_t,x_{t+1}) for any $t \geq 1$. Again, this places no restrictions on how x_t and x_{t+1} are related to one another; indeed, they may be highly correlated. Stationarity does require that the nature of any correlation between adjacent terms is the same across all time periods.

A stochastic process that is not stationary is said to be a **nonstationary process**. Since stationarity is an aspect of the underlying stochastic process and not of the available single realization, it can be very difficult to determine whether the data we have collected were generated by a stationary process. However, it is easy to spot certain sequences that are not stationary. A process with a time trend of the type covered in Section 10.5 is clearly nonstationary: at a minimum, its mean changes over time.

Sometimes, a weaker form of stationarity suffices. If $\{x_t: t = 1,2, \dots\}$ has a finite second moment, that is, $E(x_t^2) < \infty$ for all t, then the following definition applies.

COVARIANCE STATIONARY PROCESS: A stochastic process $\{x_t: t = 1,2, \dots\}$ with finite second moment $[E(x_t^2) < \infty]$ is **covariance stationary** if (i) $E(x_t)$ is constant; (ii) $\text{Var}(x_t)$ is constant; (iii) for any $t, h \geq 1$, $\text{Cov}(x_t,x_{t+h})$ depends only on h and not on t.

Covariance stationarity focuses only on the first two moments of a stochastic process: the mean and variance of the process are constant across time, and the covariance between x_t and x_{t+h} depends only on the distance between the two terms, h, and not on the location of the initial time period, t. It follows immediately that the correlation between x_t and x_{t+h} also depends only on h.

QUESTION 11.1

Suppose that $\{y_t: t = 1,2,\dots\}$ is generated by $y_t = \delta_0 + \delta_1 t + e_t$, where $\delta_1 \neq 0$, and $\{e_t: t = 1,2,\dots\}$ is an i.i.d. sequence with mean zero and variance σ_e^2. (i) Is $\{y_t\}$ covariance stationary? (ii) Is $y_t - E(y_t)$ covariance stationary?

If a stationary process has a finite second moment, then it must be covariance stationary, but the converse is certainly not true. Sometimes, to emphasize that stationarity is a stronger requirement than covariance stationarity, the former is referred to as *strict stationarity*. However, since we will not be delving into the intricacies of central limit theorems for time series processes, we will not be worried about the distinction between strict and covariance stationarity: we will call a series stationary if it satisfies either definition.

How is stationarity used in time series econometrics? On a technical level, stationarity simplifies statements of the law of large numbers and the central limit theorem, although we will not worry about formal statements. On a practical level, if we want to understand the relationship between two or more variables using regression analysis, we need to assume some sort of stability over time. If we allow the relationship between two variables (say, y_t and x_t) to change arbitrarily in each time period, then we cannot hope to learn much about how a change in one variable affects the other variable if we only have access to a single time series realization.

In stating a multiple regression model for time series data, we are assuming a certain form of stationarity in that the β_j do not change over time. Further, Assumptions TS.4 and TS.5 imply that the variance of the error process is constant over time and that the correlation between errors in two adjacent periods is equal to zero, which is clearly constant over time.

Weakly Dependent Time Series

Stationarity has to do with the joint distributions of a process as it moves through time. A very different concept is that of weak dependence, which places restrictions on how strongly related the random variables x_t and x_{t+h} can be as the time distance between them, h, gets large. The notion of weak dependence is most easily discussed for a stationary time series: loosely speaking, a stationary time series process $\{x_t: t = 1,2, \ldots\}$ is said to be **weakly dependent** if x_t and x_{t+h} are "almost independent" as h increases without bound. A similar statement holds true if the sequence is nonstationary, but then we must assume that the concept of being almost independent does not depend on the starting point, t.

The description of weak dependence given in the previous paragraph is necessarily vague. We cannot formally define weak dependence because there is no definition that covers all cases of interest. There are many specific forms of weak dependence that are formally defined, but these are well beyond the scope of this text. [See White (1984), Hamilton (1994), and Wooldridge (1994b) for advanced treatments of these concepts.]

For our purposes, an intuitive notion of the meaning of weak dependence is sufficient. Covariance stationary sequences can be characterized in terms of correlations: a covariance stationary time series is weakly dependent if the correlation between x_t and x_{t+h} goes to zero "sufficiently quickly" as $h \rightarrow \infty$. (Because of covariance stationarity, the correlation does not depend on the starting point, t.) In other words, as the variables get farther apart in time, the correlation between them becomes smaller and smaller. Covariance stationary sequences where $\text{Corr}(x_t,x_{t+h}) \rightarrow 0$ as $h \rightarrow \infty$ are said to be **asymptotically uncorrelated**. Intuitively, this is how we will usually characterize weak dependence. Technically, we need to assume that the correlation converges to zero fast enough, but we will gloss over this.

Why is weak dependence important for regression analysis? Essentially, it replaces the assumption of random sampling in implying that the law of large numbers (LLN) and the central limit theorem (CLT) hold. The most well-known central limit theorem for time series data requires stationarity and some form of weak dependence: thus, stationary, weakly dependent time series are ideal for use in multiple regression analysis. In Section 11.2, we will show how OLS can be justified quite generally by appealing to the LLN and the CLT. Time series that are not weakly dependent—examples of which we will see in Section 11.3—do not generally satisfy the CLT, which is why their use in multiple regression analysis can be tricky.

The simplest example of a weakly dependent time series is an independent, identically distributed sequence: a sequence that is independent is trivially weakly dependent. A more interesting example of a weakly dependent sequence is

$$x_t = e_t + \alpha_1 e_{t-1},\ t = 1,2,\ldots, \tag{11.1}$$

where $\{e_t: t = 0,1,\ldots\}$ is an i.i.d. sequence with zero mean and variance σ_e^2. The process $\{x_t\}$ is called a **moving average process of order one [MA(1)]**: x_t is a weighted average of e_t and e_{t-1}; in the next period, we drop e_{t-1}, and then x_{t+1} depends on e_{t+1} and e_t. Setting the coefficient on e_t to one in (11.1) is without loss of generality.

Why is an MA(1) process weakly dependent? Adjacent terms in the sequence are correlated: because $x_{t+1} = e_{t+1} + \alpha_1 e_t$, $\text{Cov}(x_t, x_{t+1}) = \alpha_1 \text{Var}(e_t) = \alpha_1 \sigma_e^2$. Since $\text{Var}(x_t) = (1 + \alpha_1^2)\sigma_e^2$, $\text{Corr}(x_t, x_{t+1}) = \alpha_1/(1 + \alpha_1^2)$. For example, if $\alpha_1 = .5$, then $\text{Corr}(x_t, x_{t+1}) = .4$. [The maximum positive correlation occurs when $\alpha_1 = 1$; in which case, $\text{Corr}(x_t, x_{t+1}) = .5$.] However, once we look at variables in the sequence that are two or more time periods apart, these variables are uncorrelated because they are independent. For example, $x_{t+2} = e_{t+2} + \alpha_1 e_{t+1}$ is independent of x_t because $\{e_t\}$ is independent across t. Due to the identical distribution assumption on the e_t, $\{x_t\}$ in (11.1) is actually stationary. Thus, an MA(1) is a stationary, weakly dependent sequence, and the law of large numbers and the central limit theorem can be applied to $\{x_t\}$.

A more popular example is the process

$$y_t = \rho_1 y_{t-1} + e_t,\ t = 1,2,\ldots. \tag{11.2}$$

The starting point in the sequence is y_0 (at $t = 0$), and $\{e_t: t = 1,2,\ldots\}$ is an i.i.d. sequence with zero mean and variance σ_e^2. We also assume that the e_t are independent of y_0 and that $\text{E}(y_0) = 0$. This is called an **autoregressive process of order one [AR(1)]**.

The crucial assumption for weak dependence of an AR(1) process is the *stability condition* $|\rho_1| < 1$. Then we say that $\{y_t\}$ is a **stable AR(1) process**.

To see that a stable AR(1) process is asymptotically uncorrelated, it is useful to assume that the process is covariance stationary. (In fact, it can generally be shown that $\{y_t\}$ is strictly stationary, but the proof is somewhat technical.) Then, we know that $\text{E}(y_t) = \text{E}(y_{t-1})$, and from (11.2) with $\rho_1 \neq 1$, this can happen only if $\text{E}(y_t) = 0$. Taking the variance of (11.2) and using the fact that e_t and y_{t-1} are independent (and therefore uncorrelated), $\text{Var}(y_t) = \rho_1^2 \text{Var}(y_{t-1}) + \text{Var}(e_t)$, and so, under covariance stationarity, we must have $\sigma_y^2 = \rho_1^2 \sigma_y^2 + \sigma_e^2$. Since $\rho_1^2 < 1$ by the stability condition, we can easily solve for σ_y^2:

$$\sigma_y^2 = \sigma_e^2/(1 - \rho_1^2). \tag{11.3}$$

Now we can find the covariance between y_t and y_{t+h} for $h \geq 1$. Using repeated substitution,

$$\begin{aligned} y_{t+h} &= \rho_1 y_{t+h-1} + e_{t+h} = \rho_1(\rho_1 y_{t+h-2} + e_{t+h-1}) + e_{t+h} \\ &= \rho_1^2 y_{t+h-2} + \rho_1 e_{t+h-1} + e_{t+h} = \dots \\ &= \rho_1^h y_t + \rho_1^{h-1} e_{t+1} + \dots + \rho_1 e_{t+h-1} + e_{t+h}. \end{aligned}$$

Since $E(y_t) = 0$ for all t, we can multiply this last equation by y_t and take expectations to obtain $\text{Cov}(y_t, y_{t+h})$. Using the fact that e_{t+j} is uncorrelated with y_t for all $j \geq 1$ gives

$$\begin{aligned} \text{Cov}(y_t, y_{t+h}) = E(y_t y_{t+h}) &= \rho_1^h E(y_t^2) + \rho_1^{h-1} E(y_t e_{t+1}) + \dots + E(y_t e_{t+h}) \\ &= \rho_1^h E(y_t^2) = \rho_1^h \sigma_y^2. \end{aligned}$$

Since σ_y is the standard deviation of both y_t and y_{t+h}, we can easily find the correlation between y_t and y_{t+h} for any $h \geq 1$:

$$\text{Corr}(y_t, y_{t+h}) = \text{Cov}(y_t, y_{t+h})/(\sigma_y \sigma_y) = \rho_1^h. \tag{11.4}$$

In particular, $\text{Corr}(y_t, y_{t+1}) = \rho_1$, so ρ_1 is the correlation coefficient between any two adjacent terms in the sequence.

Equation (11.4) is important because it shows that, while y_t and y_{t+h} are correlated for any $h \geq 1$, this correlation gets very small for large h: since $|\rho_1| < 1$, $\rho_1^h \to 0$ as $h \to \infty$. Even when ρ_1 is large—say .9, which implies a very high, positive correlation between adjacent terms—the correlation between y_t and y_{t+h} tends to zero fairly rapidly. For example, $\text{Corr}(y_t, y_{t+5}) = .591$, $\text{Corr}(y_t, y_{t+10}) = .349$, and $\text{Corr}(y_t, y_{t+20}) = .122$. If t indexes year, this means that the correlation between the outcome of two y that are twenty years apart is about .122. When ρ_1 is smaller, the correlation dies out much more quickly. (You might try $\rho_1 = .5$ to verify this.)

This analysis heuristically demonstrates that a stable AR(1) process is weakly dependent. The AR(1) model is especially important in multiple regression analysis with time series data. We will cover additional applications in Chapter 12 and the use of it for forecasting in Chapter 18.

There are many other types of weakly dependent time series, including hybrids of autoregressive and moving average processes. But the previous examples work well for our purposes.

Before ending this section, we must emphasize one point that often causes confusion in time series econometrics. A trending series, while certainly nonstationary, *can* be weakly dependent. In fact, in the simple linear time trend model in Chapter 10 [see equation (10.24)], the series $\{y_t\}$ was actually independent. A series that is stationary about its time trend, as well as weakly dependent, is often called a **trend-stationary process**. (Notice that the name is not completely descriptive because we assume weak dependence along with stationarity.) Such processes can be used in regression analysis just as in Chapter 10, *provided* appropriate time trends are included in the model.

11.2 ASYMPTOTIC PROPERTIES OF OLS

In Chapter 10, we saw some cases where the classical linear model assumptions are not satisfied for certain time series problems. In such cases, we must appeal to large sample properties of OLS, just as with cross-sectional analysis. In this section, we state the assumptions and main results that justify OLS more generally. The proofs of the theorems in this chapter are somewhat difficult and therefore omitted. See Wooldridge (1994b).

ASSUMPTION TS.1′ (LINEARITY AND WEAK DEPENDENCE)
Assumption TS.1′ is the same as TS.1, except we must also assume that $\{(\mathbf{x}_t, y_t): t = 1,2,\ldots\}$ is weakly dependent. In other words, the law of large numbers and the central limit theorem can be applied to sample averages.

The linear in parameters requirement again means that we can write the model as

$$y_t = \beta_0 + \beta_1 x_{t1} + \ldots + \beta_k x_{tk} + u_t, \tag{11.5}$$

where the β_j are the parameters to be estimated. The x_{tj} can contain lagged dependent and independent variables, provided the weak dependence assumption is met.

We have discussed the concept of weak dependence at length because it is by no means an innocuous assumption. In the next section, we will present time series processes that clearly violate the weak dependence assumption and also discuss the use of such processes in multiple regression models.

ASSUMPTION TS.2′ (ZERO CONDITIONAL MEAN)
For each t, $\mathrm{E}(u_t|\mathbf{x}_t) = 0$.

This is the most natural assumption concerning the relationship between u_t and the explanatory variables. It is much weaker than Assumption TS.2 because it puts no restrictions on how u_t is related to the explanatory variables in other time periods. We will see examples that satisfy TS.2′ shortly.

For certain purposes, it is useful to know that the following consistency result only requires u_t to have zero unconditional mean and to be uncorrelated with each x_{tj}:

$$\mathrm{E}(u_t) = 0, \ \mathrm{Cov}(x_{tj}, u_t) = 0, \ j = 1, \ldots, k. \tag{11.6}$$

We will work mostly with the zero conditional mean assumption because it leads to the most straightforward asymptotic analysis.

ASSUMPTION TS.3′ (NO PERFECT COLLINEARITY)
Same as Assumption TS.3.

THEOREM 11.1 (CONSISTENCY OF OLS)

Under TS.1′, TS.2′, and TS.3′, the OLS estimators are consistent: plim $\hat{\beta}_j = \beta_j$, $j = 0,1, \ldots, k$.

There are some key practical differences between Theorems 10.1 and 11.1. First, in Theorem 11.1, we conclude that the OLS estimators are consistent, but not necessarily unbiased. Second, in Theorem 11.1, we have weakened the sense in which the explanatory variables must be exogenous, but weak dependence is required in the underlying time series. Weak dependence is also crucial in obtaining approximate distributional results, which we cover later.

EXAMPLE 11.1
(Static Model)

Consider a static model with two explanatory variables:

$$y_t = \beta_0 + \beta_1 z_{t1} + \beta_2 z_{t2} + u_t. \tag{11.7}$$

Under weak dependence, the condition sufficient for consistency of OLS is

$$\mathrm{E}(u_t | z_{t1}, z_{t2}) = 0. \tag{11.8}$$

This rules out omitted variables that are in u_t and are correlated with either z_{t1} or z_{t2}. Also, no function of z_{t1} or z_{t2} can be correlated with u_t, and so Assumption TS.2′ rules out misspecified functional form, just as in the cross-sectional case. Other problems, such as measurement error in the variables z_{t1} or z_{t2}, can cause (11.8) to fail.

Importantly, Assumption TS.2′ *does not* rule out correlation between, say, u_{t-1} and z_{t1}. This type of correlation could arise if z_{t1} is related to past y_{t-1}, such as

$$z_{t1} = \delta_0 + \delta_1 y_{t-1} + v_t. \tag{11.9}$$

For example, z_{t1} might be a policy variable, such as monthly percentage change in the money supply, and this change depends on last month's rate of inflation (y_{t-1}). Such a mechanism generally causes z_{t1} and u_{t-1} to be correlated (as can be seen by plugging in for y_{t-1}). This kind of feedback *is* allowed under Assumption TS.2′.

EXAMPLE 11.2
(Finite Distributed Lag Model)

In the finite distributed lag model,

$$y_t = \alpha_0 + \delta_0 z_t + \delta_1 z_{t-1} + \delta_2 z_{t-2} + u_t, \tag{11.10}$$

a very natural assumption is that the expected value of u_t, given current and *all past* values of z, is zero:

$$E(u_t|z_t, z_{t-1}, z_{t-2}, z_{t-3}, \ldots) = 0. \quad \textbf{(11.11)}$$

This means that, once z_t, z_{t-1}, and z_{t-2} are included, no further lags of z affect $E(y_t|z_t, z_{t-1}, z_{t-2}, z_{t-3}, \ldots)$; if this were not true, we would put further lags into the equation. For example, y_t could be the annual percentage change in investment and z_t a measure of interest rates during year t. When we set $\mathbf{x}_t = (z_t, z_{t-1}, z_{t-2})$, Assumption TS.2′ is then satisfied: OLS will be consistent. As in the previous example, TS.2′ does not rule out feedback from y to future values of z.

The previous two examples do not necessarily require asymptotic theory because the explanatory variables *could* be strictly exogenous. The next example clearly violates the strict exogeneity assumption, and therefore we can only appeal to large sample properties of OLS.

EXAMPLE 11.3
[AR(1) Model]

Consider the AR(1) model,

$$y_t = \beta_0 + \beta_1 y_{t-1} + u_t, \quad \textbf{(11.12)}$$

where the error u_t has a zero expected value, given all past values of y:

$$E(u_t|y_{t-1}, y_{t-2}, \ldots) = 0. \quad \textbf{(11.13)}$$

Combined, these two equations imply that

$$E(y_t|y_{t-1}, y_{t-2}, \ldots) = E(y_t|y_{t-1}) = \beta_0 + \beta_1 y_{t-1}. \quad \textbf{(11.14)}$$

This result is very important. First, it means that, once y lagged one period has been controlled for, no further lags of y affect the expected value of y_t. (This is where the name "first order" originates.) Second, the relationship is assumed to be linear.

Since $\mathbf{x}_t$ contains only y_{t-1}, equation (11.13) implies that Assumption TS.2′ holds. By contrast, the strict exogeneity assumption needed for unbiasedness, Assumption TS.2, does not hold. Since the set of explanatory variables for all time periods includes all of the values on y except the last ($y_0, y_1, \ldots, y_{n-1}$), Assumption TS.2 requires that, for all t, u_t is uncorrelated with each of $y_0, y_1, \ldots, y_{n-1}$. This cannot be true. In fact, because u_t is uncorrelated with y_{t-1} under (11.13), u_t and y_t must be correlated. Therefore, a model with a lagged dependent variable cannot satisfy the strict exogeneity assumption TS.2.

For the weak dependence condition to hold, we must assume that $|\beta_1| < 1$, as we discussed in Section 11.1. If this condition holds, then Theorem 11.1 implies that the OLS estimator from the regression of y_t on y_{t-1} produces consistent estimators of β_0 and β_1. Unfortunately, $\hat{\beta}_1$ is biased, and this bias can be large if the sample size is small or if β_1 is near one. (For β_1 near one, $\hat{\beta}_1$ can have a severe downward bias.) In moderate to large samples, $\hat{\beta}_1$ should be a good estimator of β_1.

When using the standard inference procedures, we need to impose versions of the homoskedasticity and no serial correlation assumptions. These are less restrictive than their classical linear model counterparts from Chapter 10.

ASSUMPTION TS.4′ (HOMOSKEDASTICITY)
For all t, $\text{Var}(u_t|\boldsymbol{x}_t) = \sigma^2$.

ASSUMPTION TS.5′ (NO SERIAL CORRELATION)
For all $t \neq s$, $\text{E}(u_t u_s|\boldsymbol{x}_t,\boldsymbol{x}_s) = 0$.

In TS.4′, note how we condition only on the explanatory variables at time t (compare to TS.4). In TS.5′, we condition only on the explanatory variables in the time periods coinciding with u_t and u_s. As stated, this assumption is a little difficult to interpret, but it is the right condition for studying the large sample properties of OLS in a variety of time series regressions. When considering TS.5′, we often ignore the conditioning on $\boldsymbol{x}_t$ and $\boldsymbol{x}_s$, and we think about whether u_t and u_s are uncorrelated, for all $t \neq s$.

Serial correlation is often a problem in static and finite distributed lag regression models: nothing guarantees that the unobservables u_t are uncorrelated over time. Importantly, Assumption TS.5′ *does* hold in the AR(1) model stated in equations (11.12) and (11.13). Since the explanatory variable at time t is y_{t-1}, we must show that $\text{E}(u_t u_s|y_{t-1},y_{s-1}) = 0$ for all $t \neq s$. To see this, suppose that $s < t$. (The other case follows by symmetry.) Then, since $u_s = y_s - \beta_0 - \beta_1 y_{s-1}$, u_s is a function of y dated before time t. But by (11.13), $\text{E}(u_t|u_s,y_{t-1},y_{s-1}) = 0$, and then the law of iterated expectations (see Appendix B) implies that $\text{E}(u_t u_s|y_{t-1},y_{s-1}) = 0$. This is very important: as long as only one lag belongs in (11.12), the errors must be serially uncorrelated. We will discuss this feature of dynamic models more generally in Section 11.4.

We now obtain an asymptotic result that is practically identical to the cross-sectional case.

THEOREM 11.2 (ASYMPTOTIC NORMALITY OF OLS)
Under TS.1′ through TS.5′, the OLS estimators are asymptotically normally distributed. Further, the usual OLS standard errors, *t* statistics, *F* statistics, and *LM* statistics are asymptotically valid.

This theorem provides additional justification for at least some of the examples estimated in Chapter 10: even if the classical linear model assumptions do not hold, OLS is still consistent, and the usual inference procedures are valid. Of course, this hinges on TS.1′ through TS.5′ being true. In the next section, we discuss ways in which the weak dependence assumption can fail. The problems of serial correlation and heteroskedasticity are treated in Chapter 12.

EXAMPLE 11.4
(Efficient Markets Hypothesis)

We can use asymptotic analysis to test a version of the *efficient markets hypothesis* (EMH). Let y_t be the weekly percentage return (from Wednesday close to Wednesday close) on the *New York Stock Exchange* composite index. A strict form of the efficient markets hypothesis states that information observable to the market prior to week t should not help to predict the return during week t. If we use only past information on y, the EMH is stated as

$$\mathrm{E}(y_t|y_{t-1},y_{t-2},\ldots) = \mathrm{E}(y_t). \quad \textbf{(11.15)}$$

If (11.15) is false, then we could use information on past weekly returns to predict the current return. The EMH presumes that such investment opportunities will be noticed and will disappear almost instantaneously.

One simple way to test (11.15) is to specify the AR(1) model in (11.12) as the alternative model. Then, the null hypothesis is easily stated as H_0: $\beta_1 = 0$. Under the null hypothesis, Assumption TS.2′ is true by (11.15), and, as we discussed earlier, serial correlation is not an issue. The homoskedasticity assumption is $\mathrm{Var}(y_t|y_{t-1}) = \mathrm{Var}(y_t) = \sigma^2$, which we just assume is true for now. Under the null hypothesis, stock returns are serially uncorrelated, so we can safely assume that they are weakly dependent. Then, Theorem 11.2 says we can use the usual OLS t statistic for $\hat{\beta}_1$ to test H_0: $\beta_1 = 0$ against H_1: $\beta_1 \neq 0$.

The weekly returns in NYSE.RAW are computed using data from January 1976 through March 1989. In the rare case that Wednesday was a holiday, the close at the next trading day was used. The average weekly return over this period was .196 in percent form, with the largest weekly return being 8.45% and the smallest being −15.32% (during the stock market crash of October 1987). Estimation of the AR(1) model gives

$$\begin{gathered}\widehat{return}_t = \underset{(.081)}{.180} + \underset{(.038)}{.059}\, return_{t-1} \\ n = 689,\ R^2 = .0035,\ \bar{R}^2 = .0020.\end{gathered} \quad \textbf{(11.16)}$$

The t statistic for the coefficient on $return_{t-1}$ is about 1.55, and so H_0: $\beta_1 = 0$ cannot be rejected against the two-sided alternative, even at the 10% significance level. The estimate does suggest a slight positive correlation in the NYSE return from one week to the next, but it is not strong enough to warrant rejection of the efficient markets hypothesis.

In the previous example, using an AR(1) model to test the EMH might not detect correlation between weekly returns that are more than one week apart. It is easy to estimate models with more than one lag. For example, an *autoregressive model of order two*, or AR(2) model, is

$$\begin{gathered}y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + u_t \\ \mathrm{E}(u_t|y_{t-1},y_{t-2},\ldots) = 0.\end{gathered} \quad \textbf{(11.17)}$$

There are stability conditions on β_1 and β_2 that are needed to ensure that the AR(2) process is weakly dependent, but this is not an issue here because the null hypothesis states that the EMH holds:

$$H_0: \beta_1 = \beta_2 = 0. \quad \textbf{(11.18)}$$

If we add the homoskedasticity assumption $\text{Var}(u_t|y_{t-1},y_{t-2}) = \sigma^2$, we can use a standard F statistic to test (11.18). If we estimate an AR(2) model for $return_t$, we obtain

$$\underset{}{\widehat{return}_t} = \underset{(.081)}{.186} + \underset{(.038)}{.060}\, return_{t-1} - \underset{(.038)}{.038}\, return_{t-2}$$

$$n = 688,\ R^2 = .0048,\ \bar{R}^2 = .0019$$

(where we lose one more observation because of the additional lag in the equation). The two lags are individually insignificant at the 10% level. They are also jointly insignificant: using $R^2 = .0048$, the F statistic is approximately $F = 1.65$; the p-value for this F statistic (with 2 and 685 degrees of freedom) is about .193. Thus, we do no reject (11.18) at even the 15% significance level.

EXAMPLE 11.5
(Expectations Augmented Phillips Curve)

A linear version of the *expectations augmented Phillips curve* can be written as

$$inf_t - inf_t^e = \beta_1(unem_t - \mu_0) + e_t,$$

where μ_0 is the *natural rate of unemployment* and inf_t^e is the *expected* rate of inflation formed in year $t - 1$. This model assumes that the natural rate is constant, something that macroeconomists question. The difference between actual unemployment and the natural rate is called *cyclical unemployment*, while the difference between actual and expected inflation is called *unanticipated inflation*. The error term, e_t, is called a *supply shock* by macroeconomists. If there is a tradeoff between unanticipated inflation and cyclical unemployment, then $\beta_1 < 0$. [For a detailed discussion of the expectations augmented Phillips curve, see Mankiw (1994, Section 11.2).]

To complete this model, we need to make an assumption about inflationary expectations. Under *adaptive expectations*, the expected value of current inflation depends on recently observed inflation. A particularly simple formulation is that expected inflation this year is last year's inflation: $inf_t^e = inf_{t-1}$. (See Section 18.1 for an alternative formulation of adaptive expectations.) Under this assumption, we can write

$$inf_t - inf_{t-1} = \beta_0 + \beta_1 unem_t + e_t$$

or

$$\Delta inf_t = \beta_0 + \beta_1 unem_t + e_t,$$

where $\Delta inf_t = inf_t - inf_{t-1}$ and $\beta_0 = -\beta_1\mu_0$. (β_0 is expected to be positive, since $\beta_1 < 0$ and $\mu_0 > 0$.) Therefore, under adaptive expectations, the expectations augmented Phillips curve relates the *change* in inflation to the level of unemployment and a supply shock, e_t. If e_t is uncorrelated with $unem_t$, as is typically assumed, then we can consistently estimate

β_0 and β_1 by OLS. (We do not have to assume that, say, future unemployment rates are unaffected by the current supply shock.) We assume that TS.1′ through TS.5′ hold. The estimated equation is

$$\Delta \widehat{inf}_t = 3.03 - .543\ unem_t$$
$$(1.38)\quad (.230)$$
$$n = 48,\ R^2 = .108,\ \bar{R}^2 = .088. \tag{11.19}$$

The tradeoff between cyclical unemployment and unanticipated inflation is pronounced in equation (11.19): a one-point increase in *unem* lowers unanticipated inflation by over one-half of a point. The effect is statistically significant (two-sided p-value $\approx$ .023). We can contrast this with the static Phillips curve in Example 10.1, where we found a slightly positive relationship between inflation and unemployment.

Because we can write the natural rate as $\mu_0 = \beta_0/(-\beta_1)$, we can use (11.19) to obtain our own estimate of the natural rate: $\hat{\mu}_0 = \hat{\beta}_0/(-\hat{\beta}_1) = 3.03/.543 \approx 5.58$. Thus, we estimate the natural rate to be about 5.6, which is well within the range suggested by macroeconomists: historically, 5 to 6% is a common range cited for the natural rate of unemployment. It is possible to obtain an approximate standard error for this estimate, but the methods are beyond the scope of this text. [See, for example, Davidson and MacKinnon (1993).]

Under Assumptions TS.1′ through TS.5′, we can show that the OLS estimators are asymptotically efficient in the class of estimators described in Theorem 5.3, but we replace the cross-sectional observation index i with the time series index t. Finally, models with trending explanatory variables can satisfy Assumptions TS.1′ through TS.5′, provided they are trend stationary. As long as time trends are included in the equations when needed, the usual inference procedures are asymptotically valid.

QUESTION 11.2

Suppose that expectations are formed as $inf_t^e = (1/2)inf_{t-1} + (1/2)inf_{t-2}$. What regression would you run to estimate the expectations augmented Phillips curve?

11.3 USING HIGHLY PERSISTENT TIME SERIES IN REGRESSION ANALYSIS

The previous section shows that, provided the time series we use are weakly dependent, usual OLS inference procedures are valid under assumptions weaker than the classical linear model assumptions. Unfortunately, many economic time series cannot be characterized by weak dependence. Using time series with strong dependence in regression analysis poses no problem, *if* the CLM assumptions in Chapter 10 hold. But the usual inference procedures are very susceptible to violation of these assumptions when the data are not weakly dependent, because then we cannot appeal to the law of large numbers and the central limit theorem. In this section, we provide some examples of **highly**

persistent (or **strongly dependent**) time series and show how they can be transformed for use in regression analysis.

Highly Persistent Time Series

In the simple AR(1) model (11.2), the assumption $|\rho_1| < 1$ is crucial for the series to be weakly dependent. It turns out that many economic time series are better characterized by the AR(1) model with $\rho_1 = 1$. In this case, we can write

$$y_t = y_{t-1} + e_t, t = 1,2, \ldots, \qquad \textbf{(11.20)}$$

where we again assume that $\{e_t: t = 1,2,\ldots\}$ is independent and identically distributed with mean zero and variance σ_e^2. We assume that the initial value, y_0, is independent of e_t for all $t \geq 1$.

The process in (11.20) is called a **random walk**. The name comes from the fact that y at time t is obtained by starting at the previous value, y_{t-1}, and adding a zero mean random variable that is independent of y_{t-1}. Sometimes, a random walk is defined differently by assuming different properties of the innovations, e_t (such as lack of correlation rather than independence), but the current definition suffices for our purposes.

First, we find the expected value of y_t. This is most easily done by using repeated substitution to get

$$y_t = e_t + e_{t-1} + \ldots + e_1 + y_0.$$

Taking the expected value of both sides gives

$$\begin{aligned} \mathrm{E}(y_t) &= \mathrm{E}(e_t) + \mathrm{E}(e_{t-1}) + \ldots + \mathrm{E}(e_1) + \mathrm{E}(y_0) \\ &= \mathrm{E}(y_0), \text{ for all } t \geq 1. \end{aligned}$$

Therefore, the expected value of a random walk does *not* depend on t. A popular assumption is that $y_0 = 0$—the process begins at zero at time zero—in which case, $\mathrm{E}(y_t) = 0$ for all t.

By contrast, the variance of a random walk does change with t. To compute the variance of a random walk, for simplicity we assume that y_0 is nonrandom so that $\mathrm{Var}(y_0) = 0$; this does not affect any important conclusions. Then, by the i.i.d. assumption for $\{e_t\}$,

$$\mathrm{Var}(y_t) = \mathrm{Var}(e_t) + \mathrm{Var}(e_{t-1}) + \ldots + \mathrm{Var}(e_1) = \sigma_e^2 t. \qquad \textbf{(11.21)}$$

In other words, the variance of a random walk increases as a linear function of time. This shows that the process cannot be stationary.

Even more importantly, a random walk displays highly persistent behavior in the sense that the value of y today is significant for determining the value of y in the very distant future. To see this, write for h periods hence,

$$y_{t+h} = e_{t+h} + e_{t+h-1} + \ldots + e_{t+1} + y_t.$$

Now, suppose at time t, we want to compute the expected value of y_{t+h} given the current value y_t. Since the expected value of e_{t+j}, given y_t, is zero for all $j \geq 1$, we have

$$\mathrm{E}(y_{t+h}|y_t) = y_t, \text{ for all } h \geq 1. \tag{11.22}$$

This means that, no matter how far in the future we look, our best prediction of y_{t+h} is today's value, y_t. We can contrast this with the stable AR(1) case, where a similar argument can be used to show that

$$\mathrm{E}(y_{t+h}|y_t) = \rho_1^h y_t, \text{ for all } h \geq 1.$$

Under stability, $|\rho_1| < 1$, and so $\mathrm{E}(y_{t+h}|y_t)$ approaches zero as $h \rightarrow \infty$: the value of y_t becomes less and less important, and $\mathrm{E}(y_{t+h}|y_t)$ gets closer and closer to the unconditional expected value, $\mathrm{E}(y_t) = 0$.

When $h = 1$, equation (11.22) is reminiscent of the adaptive expectations assumption we used for the inflation rate in Example 11.5: if inflation follows a random walk, then the expected value of inf_t, given past values of inflation, is simply inf_{t-1}. Thus, a random walk model for inflation justifies the use of adaptive expectations.

We can also see that the correlation between y_t and y_{t+h} is close to one for large t when $\{y_t\}$ follows a random walk. If $\mathrm{Var}(y_0) = 0$, it can be shown that

$$\mathrm{Corr}(y_t, y_{t+h}) = \sqrt{t/(t+h)}.$$

Thus, the correlation depends on the starting point, t (so that $\{y_t\}$ is not covariance stationary). Further, for fixed t, the correlation tends to zero as $h \rightarrow 0$, but it does not do so very quickly. In fact, the larger t is, the more slowly the correlation tends to zero as h gets large. If we choose h to be something large—say, $h = 100$—we can always choose a large enough t such that the correlation between y_t and y_{t+h} is arbitrarily close to one. (If $h = 100$ and we want the correlation to be greater than .95, then $t > 1{,}000$ does the trick.) Therefore, a random walk does not satisfy the requirement of an asymptotically uncorrelated sequence.

Figure 11.1 plots two realizations of a random walk with initial value $y_0 = 0$ and $e_t \sim \mathrm{Normal}(0,1)$. Generally, it is not easy to look at a time series plot and to determine whether or not it is a random walk. Next, we will discuss an informal method for making the distinction between weakly and highly dependent sequences; we will study formal statistical tests in Chapter 18.

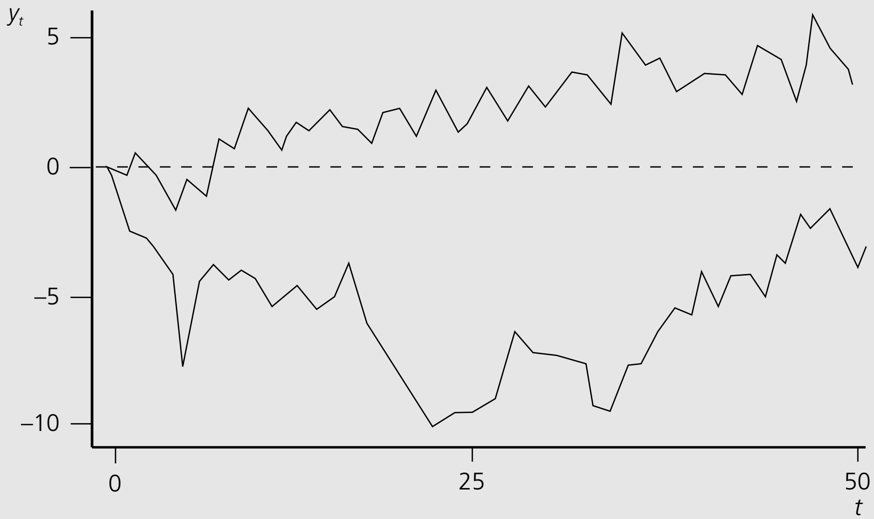
A series that is generally thought to be well-characterized by a random walk is the three-month, T-bill rate. Annual data are plotted in Figure 11.2 for the years 1948 through 1996.

A random walk is a special case of what is known as a **unit root process**. The name comes from the fact that $\rho_1 = 1$ in the AR(1) model. A more general class of unit root processes is generated as in (11.20), but $\{e_t\}$ is now allowed to be a general, weakly dependent series. [For example, $\{e_t\}$ could itself follow an MA(1) or a stable AR(1) process.] When $\{e_t\}$ is not an i.i.d. sequence, the properties of the random walk we derived earlier no longer hold. But the key feature of $\{y_t\}$ is preserved: the value of y today is highly correlated with y even in the distant future.

From a policy perspective, it is often important to know whether an economic time series is highly persistent or not. Consider the case of gross domestic product in the United States. If GDP is asymptotically uncorrelated, then the level of GDP in the coming year is at best weakly related to what GDP was, say, thirty years ago. This means a policy that affected GDP long ago has very little lasting impact. On the other hand, if

Figure 11.1

Two realizations of the random walk $y_t = y_{t-1} + e_t$, with $y_0 = 0$, $e_t \sim \text{Normal}(0,1)$, and $n = 50$.



GDP is strongly dependent, then next year's GDP can be highly correlated with the GDP from many years ago. Then, we should recognize that a policy which causes a discrete change in GDP can have long-lasting effects.

It is extremely important not to confuse trending and highly persistent behaviors. A series can be trending but not highly persistent, as we saw in Chapter 10. Further, factors such as interest rates, inflation rates, and unemployment rates are thought by many to be highly persistent, but they have no obvious upward or downward trend. However, it is often the case that a highly persistent series also contains a clear trend. One model that leads to this behavior is the **random walk with drift**:

$$y_t = \alpha_0 + y_{t-1} + e_t, \; t = 1,2, \ldots, \qquad \textbf{(11.23)}$$

where $\{e_t: t = 1,2, \ldots\}$ and y_0 satisfy the same properties as in the random walk model. What is new is the parameter α_0, which is called the *drift term*. Essentially, to generate y_t, the constant α_0 is added along with the random noise e_t to the previous value y_{t-1}. We can show that the expected value of y_t follows a linear time trend by using repeated substitution:

$$y_t = \alpha_0 t + e_t + e_{t-1} + \ldots + e_1 + y_0.$$

Therefore, if $y_0 = 0$, $\text{E}(y_t) = \alpha_0 t$: the expected value of y_t is growing over time if $\alpha_0 > 0$ and shrinking over time if $\alpha_0 < 0$. By reasoning as we did in the pure random walk case, we can show that $\text{E}(y_{t+h}|y_t) = \alpha_0 h + y_t$, and so the best prediction of y_{t+h} at time t is y_t plus the drift $\alpha_0 h$. The variance of y_t is the same as it was in the pure random walk case.

Figure 11.2

The U.S. three-month T-bill rate, for the years 1948–1996.

Figure 11.3 contains a realization of a random walk with drift, where $n = 50$, $y_0 = 0$, $\alpha_0 = 2$, and the e_t are Normal(0,9) random variables. As can be seen from this graph, y_t tends to grow over time, but the series does not regularly return to the trend line.

A random walk with drift is another example of a unit root process, because it is the special case $\rho_1 = 1$ in an AR(1) model with an intercept:

$$y_t = \alpha_0 + \rho_1 y_{t-1} + e_t.$$

When $\rho_1 = 1$ and $\{e_t\}$ is any weakly dependent process, we obtain a whole class of highly persistent time series processes that also have linearly trending means.

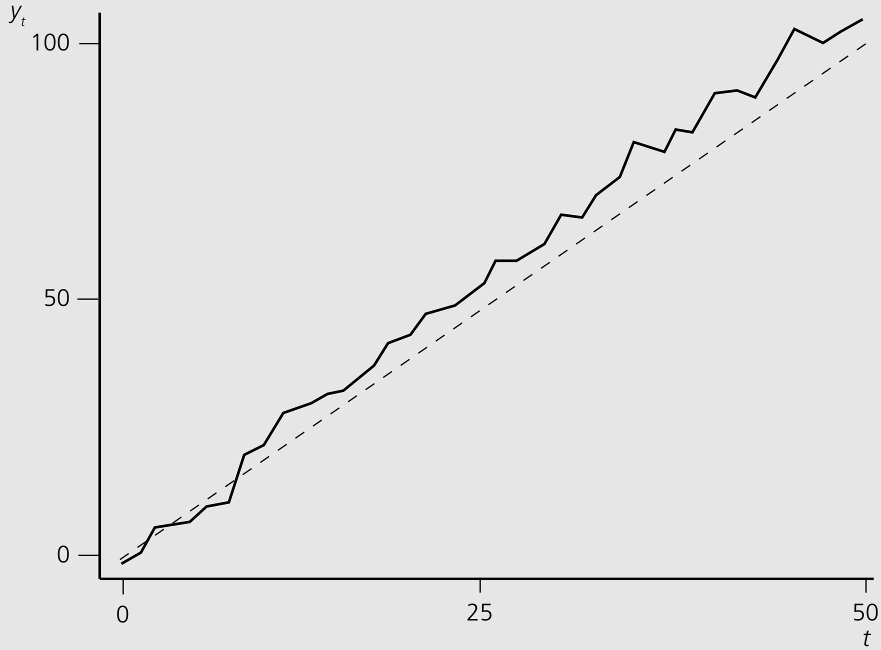
Transformations on Highly Persistent Time Series

Using time series with strong persistence of the type displayed by a unit root process in a regression equation can lead to very misleading results if the CLM assumptions are violated. We will study the spurious regression problem in more detail in Chapter 18, but for now we must be aware of potential problems. Fortunately, simple transformations are available that render a unit root process weakly dependent.

Weakly dependent processes are said to be **integrated of order zero, [I(0)]**. Practically, this means that nothing needs to be done to such series before using them in regression analysis: averages of such sequences already satisfy the standard limit the-

Figure 11.3

A realization of the random walk with drift, $y_t = 2 + y_{t-1} + e_t$, with $y_0 = 0$, $e_t \sim$ Normal(0,9), and $n = 50$. The dashed line is the expected value of y_t, $E(y_t) = 2t$.



orems. Unit root processes, such as a random walk (with or without drift), are said to be **integrated of order zero**, or **I(0)**. This means that the **first difference** of the process is weakly dependent (and often stationary).

This is simple to see for a random walk. With $\{y_t\}$ generated as in (11.20) for $t = 1,2, \ldots,$

$$\Delta y_t = y_t - y_{t-1} = e_t, t = 2,3, \ldots; \quad \textbf{(11.24)}$$

therefore, the first-differenced series $\{\Delta y_t: t = 2,3, \ldots\}$ is actually an i.i.d. sequence. More generally, if $\{y_t\}$ is generated by (11.24) where $\{e_t\}$ is any weakly dependent process, then $\{\Delta y_t\}$ is weakly dependent. Thus, when we suspect processes are integrated of order one, we often first difference in order to use them in regression analysis; we will see some examples later.

Many time series y_t that are strictly positive are such that $\log(y_t)$ is integrated of order one. In this case, we can use the first difference in the logs, $\Delta\log(y_t) = \log(y_t) - \log(y_{t-1})$, in regression analysis. Alternatively, since

$$\Delta\log(y_t) \approx (y_t - y_{t-1})/y_{t-1}, \quad \textbf{(11.25)}$$

we can use the proportionate or percentage change in y_t directly; this is what we did in Example 11.4 where, rather than stating the efficient markets hypothesis in terms of the stock price, p_t, we used the weekly percentage change, $return_t = 100[(p_t - p_{t-1})/p_{t-1}]$.

Differencing time series before using them in regression analysis has another benefit: it removes any linear time trend. This is easily seen by writing a linearly trending variable as

$$y_t = \gamma_0 + \gamma_1 t + v_t,$$

where v_t has a zero mean. Then $\Delta y_t = \gamma_1 + \Delta v_t$, and so $\text{E}(\Delta y_t) = \gamma_1 + \text{E}(\Delta v_t) = \gamma_1$. In other words, $\text{E}(\Delta y_t)$ is constant. The same argument works for $\Delta\log(y_t)$ when $\log(y_t)$ follows a linear time trend. Therefore, rather than including a time trend in a regression, we can instead difference those variables that show obvious trends.

Deciding Whether a Time Series Is I(1)

Determining whether a particular time series realization is the outcome of an I(1) versus an I(0) process can be quite difficult. Statistical tests can be used for this purpose, but these are more advanced; we provide an introductory treatment in Chapter 18.

There are informal methods that provide useful guidance about whether a time series process is roughly characterized by weak dependence. A very simple tool is motivated by the AR(1) model: if $|\rho_1| < 1$, then the process is I(0), but it is I(1) if $\rho_1 = 1$. Earlier, we showed that, when the AR(1) process is stable, $\rho_1 = \text{Corr}(y_t, y_{t-1})$. Therefore, we can estimate ρ_1 from the sample correlation between y_t and y_{t-1}. This sample correlation coefficient is called the *first order autocorrelation* of $\{y_t\}$; we denote this by $\hat{\rho}_1$. By applying the law of large numbers, $\hat{\rho}_1$ can be shown to be consistent for ρ_1 *provided* $|\rho_1| < 1$. (However, $\hat{\rho}_1$ is not an unbiased estimator of ρ_1.)

We can use the value of $\hat{\rho}_1$ to help decide whether the process is I(1) or I(0). Unfortunately, because $\hat{\rho}_1$ is an estimate, we can never know for sure whether $\rho_1 < 1$. Ideally, we could compute a confidence interval for ρ_1 to see if it excludes the value $\rho_1 = 1$, but this turns out to be rather difficult: the sampling distributions of the estimator of $\hat{\rho}_1$ are extremely different when ρ_1 is close to one and when ρ_1 is much less than one. (In fact, when ρ_1 is close to one, $\hat{\rho}_1$ can have a severe downward bias.)

In Chapter 18, we will show how to test H_0: $\rho_1 = 1$ against H_0: $\rho_1 < 1$. For now, we can only use $\hat{\rho}_1$ as a rough guide for determining whether a series needs to be differenced. No hard and fast rule exists for making this choice. Most economists think that differencing is warranted if $\hat{\rho}_1 > .9$; some would difference when $\hat{\rho}_1 > .8$.

EXAMPLE 11.6
(Fertility Equation)

In Example 10.4, we explained the general fertility rate, *gfr*, in terms of the value of the personal exemption, *pe*. The first order autocorrelations for these series are very large: $\hat{\rho}_1 = .977$ for *gfr* and $\hat{\rho}_1 = .964$ for *pe*. These are suggestive of unit root behavior, and they raise questions about the use of the usual OLS *t* statistics in Chapter 10. We now estimate the equations using the first differences (and dropping the dummy variables for simplicity):

$$\Delta\widehat{gfr} = -.785 - .043\ \Delta pe \atop (.502)\quad (.028)$$
$$n = 71,\ R^2 = .032,\ \bar{R}^2 = .018. \tag{11.26}$$

Now, an increase in *pe* is estimated to lower *gfr* contemporaneously, although the estimate is not statistically different from zero at the 5% level. This gives very different results than when we estimated the model in levels, and it casts doubt on our earlier analysis.

If we add two lags of Δpe, things improve:

$$\Delta\widehat{gfr} = -.964 - .036\ \Delta pe - .014\ \Delta pe_{-1} + .110\ \Delta pe_{-2} \atop (.468)\quad (.027)\qquad\quad (.028)\qquad\qquad (.027)$$
$$n = 69,\ R^2 = .233,\ \bar{R}^2 = .197. \tag{11.27}$$

Even though Δpe and Δpe_{-1} have negative coefficients, their coefficients are small and jointly insignificant (*p*-value = .28). The second lag is very significant and indicates a positive relationship between changes in *pe* and subsequent changes in *gfr* two years hence. This makes more sense than having a contemporaneous effect. See Exercise 11.12 for further analysis of the equation in first differences.

When the series in question has an obvious upward or downward trend, it makes more sense to obtain the first order autocorrelation after detrending. If the data are not detrended, the autoregressive correlation tends to be overestimated, which biases toward finding a unit root in a trending process.

EXAMPLE 11.7
(Wages and Productivity)

The variable *hrwage* is average hourly wage in the U.S. economy, and *outphr* is output per hour. One way to estimate the elasticity of hourly wage with respect to output per hour is to estimate the equation,

$$\log(hrwage_t) = \beta_0 + \beta_1 \log(outphr_t) + \beta_2 t + u_t,$$

where the time trend is included because log(*hrwage*) and log($outphr_t$) both display clear, upward, linear trends. Using the data in EARNS.RAW for the years 1947 through 1987, we obtain

$$\log(\widehat{hrwage_t}) = -5.33 + 1.64\ \log(outphr_t) - .018\ t \atop (0.37)\quad (0.09)\qquad\qquad\qquad (.002)$$
$$n = 41,\ R^2 = .971,\ \bar{R}^2 = .970. \tag{11.28}$$

(We have reported the usual goodness-of-fit measures here; it would be better to report those based on the detrended dependent variable, as in Section 10.5.) The estimated elasticity seems too large: a 1% increase in productivity increases real wages by about 1.64%.

Because the standard error is so small, the 95% confidence interval easily excludes a unit elasticity. U.S. workers would probably have trouble believing that their wages increase by more than 1.5% for every 1% increase in productivity.

The regression results in (11.28) must be viewed with caution. Even after linearly detrending log(*hrwage*), the first order autocorrelation is .967, and for detrended log(*outphr*), $\hat{\rho}_1 = .945$. These suggest that both series have unit roots, so we reestimate the equation in first differences (and we no longer need a time trend):

$$\begin{aligned} \Delta\log(\widehat{hrwage_t}) &= -.0036 + .809\ \Delta\log(outphr) \\ &\quad\ \ (.0042)\quad (.173) \\ n &= 40,\ R^2 = .364,\ \bar{R}^2 = .348. \end{aligned} \tag{11.29}$$

Now, a 1% increase in productivity is estimated to increase real wages by about .81%, and the estimate is not statistically different from one. The adjusted *R*-squared shows that the growth in output explains about 35% of the growth in real wages. See Exercise 11.9 for a simple distributed lag version of the model in first differences.

In the previous two examples, both the dependent and independent variables appear to have unit roots. In other cases, we might have a mixture of processes with unit roots and those that are weakly dependent (though possibly trending). An example is given in Exercise 11.8.

11.4 DYNAMICALLY COMPLETE MODELS AND THE ABSENCE OF SERIAL CORRELATION

In the AR(1) model (11.12), we showed that, under assumption (11.13), the errors $\{u_t\}$ *must* be **serially uncorrelated** in the sense that Assumption TS.5′ is satisfied: assuming that no serial correlation exists is practically the same thing as assuming that only one lag of y appears in $E(y_t|y_{t-1},y_{t-2}, \ldots)$.

Can we make a similar statement for other regression models? The answer is yes. Consider the simple static regression model

$$y_t = \beta_0 + \beta_1 z_t + u_t, \tag{11.30}$$

where y_t and z_t are contemporaneously dated. For consistency of OLS, we only need $E(u_t|z_t) = 0$. Generally, the $\{u_t\}$ will be serially correlated. However, if we *assume* that

$$E(u_t|z_t,y_{t-1},z_{t-1}, \ldots) = 0, \tag{11.31}$$

then (as we will show generally later) Assumption TS.5′ holds. In particular, the $\{u_t\}$ are serially uncorrelated.

To gain insight into the meaning of (11.31), we can write (11.30) and (11.31) equivalently as

$$E(y_t|z_t,y_{t-1},z_{t-1},\ldots) = E(y_t|z_t) = \beta_0 + \beta_1 z_t, \tag{11.32}$$

where the first equality is the one of current interest. It says that, once z_t has been controlled for, no lags of either y or z help to explain current y. This is a strong requirement; if it is false, then we can expect the errors to be serially correlated.

Next, consider a finite distributed lag model with two lags:

$$y_t = \beta_0 + \beta_1 z_t + \beta_2 z_{t-1} + \beta_3 z_{t-2} + u_t. \tag{11.33}$$

Since we are hoping to capture the lagged effects that z has on y, we would naturally assume that (11.33) captures the *distributed lag dynamics*:

$$E(y_t|z_t,z_{t-1},z_{t-2},z_{t-3},\ldots) = E(y_t|z_t,z_{t-1},z_{t-2}); \tag{11.34}$$

that is, at most two lags of z matter. If (11.31) holds, we can make further statements: once we have controlled for z and its two lags, no lags of y or additional lags of z affect current y:

$$E(y_t|z_t,y_{t-1},z_{t-1},\ldots) = E(y_t|z_t,z_{t-1},z_{t-2}). \tag{11.35}$$

Equation (11.35) is more likely than (11.32), but it still rules out lagged y affecting current y.

Next, consider a model with one lag of both y and z:

$$y_t = \beta_0 + \beta_1 z_t + \beta_2 y_{t-1} + \beta_3 z_{t-1} + u_t.$$

Since this model includes a lagged dependent variable, (11.31) is a natural assumption, as it implies that

$$E(y_t|z_t,y_{t-1},z_{t-1},y_{t-2}\ldots) = E(y_t|z_t,y_{t-1},z_{t-1});$$

in other words, once z_t, y_{t-1}, and z_{t-1} have been controlled for, no further lags of either y or z affect current y.

In the general model

$$y_t = \beta_0 + \beta_1 x_{t1} + \ldots + \beta_k x_{tk} + u_t, \tag{11.36}$$

where the explanatory variables $\boldsymbol{x}_t = (x_{t1}, \ldots, x_{tk})$ may or may not contain lags of y or z, (11.31) becomes

$$E(u_t|\boldsymbol{x}_t,y_{t-1},\boldsymbol{x}_{t-1},\ldots) = 0. \tag{11.37}$$

Written in terms of y_t,

$$E(y_t|\boldsymbol{x}_t,y_{t-1},\boldsymbol{x}_{t-1},\ldots) = E(y_t|\boldsymbol{x}_t). \tag{11.38}$$

In words, whatever is in $\boldsymbol{x}_t$, enough lags have been included so that further lags of y and the explanatory variables do not matter for explaining y_t. When this condition holds, we have a **dynamically complete model**. As we saw earlier, dynamic completeness can be a very strong assumption for static and finite distributed lag models.

Once we start putting lagged y as explanatory variables, we often think that the model should be dynamically complete. We will touch on some exceptions to this practice in Chapter 18.

Since (11.37) is equivalent to

$$\mathrm{E}(u_t|\boldsymbol{x}_t,u_{t-1},\boldsymbol{x}_{t-1},u_{t-2}, \ldots) = 0, \tag{11.39}$$

we can show that a dynamically complete model *must* satisfy Assumption TS.5′. (This derivation is not crucial and can be skipped without loss of continuity.) For concreteness, take $s < t$. Then, by the law of iterated expectations (see Appendix B),

$$\begin{aligned}\mathrm{E}(u_tu_s|\boldsymbol{x}_t,\boldsymbol{x}_s) &= \mathrm{E}[\mathrm{E}(u_tu_s|\boldsymbol{x}_t,\boldsymbol{x}_s,u_s)|\boldsymbol{x}_t,\boldsymbol{x}_s] \\ &= \mathrm{E}[u_s\mathrm{E}(u_t|\boldsymbol{x}_t,\boldsymbol{x}_s,u_s)|\boldsymbol{x}_t,\boldsymbol{x}_s],\end{aligned}$$

where the second equality follows from $\mathrm{E}(u_tu_s|\boldsymbol{x}_t,\boldsymbol{x}_s,u_s) = u_s\mathrm{E}(u_t|\boldsymbol{x}_t,\boldsymbol{x}_s,u_s)$. Now, since $s < t$, $(\boldsymbol{x}_t,\boldsymbol{x}_s,u_s)$ is a subset of the conditioning set in (11.39). Therefore, (11.39) implies that $\mathrm{E}(u_t|\boldsymbol{x}_t,\boldsymbol{x}_s,u_s) = 0$, and so

$$\mathrm{E}(u_tu_s|\boldsymbol{x}_t,\boldsymbol{x}_s) = \mathrm{E}(u_s \cdot 0|\boldsymbol{x}_t,\boldsymbol{x}_s) = 0,$$

which says that Assumption TS.5′ holds.

Since specifying a dynamically complete model means that there is no serial correlation, does it follow that all models should be dynamically complete? As we will see in Chapter 18, for forecasting purposes, the answer is yes. Some think that all models should be dynamically complete and that serial correlation in the errors of a model is a sign of misspecification. This stance is too rigid. Sometimes, we really are interested in a static model (such as a Phillips curve) or a finite distributed lag model (such as measuring the long-run percentage change in wages given a 1% increase in productivity). In the next chapter, we will show how to detect and correct for serial correlation in such models.

QUESTION 11.3

If (11.33) holds where $u_t = e_t + \alpha_1 e_{t-1}$ and where $\{e_t\}$ is an i.i.d. sequence with mean zero and variance σ_e^2, can equation (11.33) be dynamically complete?

EXAMPLE 11.8

(Fertility Equation)

In equation (11.27), we estimated a distributed lag model for Δgfr on Δpe, allowing for two lags of Δpe. For this model to be dynamically complete in the sense of (11.38), neither lags of Δgfr nor further lags of Δpe should appear in the equation. We can easily see that this is false by adding Δgfr_{-1}: the coefficient estimate is .300, and its t statistic is 2.84. Thus, the model is not dynamically complete in the sense of (11.38).

What should we make of this? We will postpone an interpretation of general models with lagged dependent variables until Chapter 18. But the fact that (11.27) is not dynamically complete suggests that there may be serial correlation in the errors. We will see how to test and correct for this in Chapter 12.

11.5 THE HOMOSKEDASTICITY ASSUMPTION FOR TIME SERIES MODELS

The homoskedasticity assumption for time series regressions, particularly TS.4′, looks very similar to that for cross-sectional regressions. However, since $\mathbf{x}_t$ can contain lagged y as well as lagged explanatory variables, we briefly discuss the meaning of the homoskedasticity assumption for different time series regressions.

In the simple static model, say

$$y_t = \beta_0 + \beta_1 z_t + u_t, \tag{11.37}$$

Assumption TS.4′ requires that

$$\text{Var}(u_t|z_t) = \sigma^2.$$

Therefore, even though $\text{E}(y_t|z_t)$ is a linear function of z_t, $\text{Var}(y_t|z_t)$ must be constant. This is pretty straightforward.

In Example 11.4, we saw that, for the AR(1) model (11.12), the homoskedasticity assumption is

$$\text{Var}(u_t|y_{t-1}) = \text{Var}(y_t|y_{t-1}) = \sigma^2;$$

even though $\text{E}(y_t|y_{t-1})$ depends on y_{t-1}, $\text{Var}(y_t|y_{t-1})$ does not. Thus, the variation in the distribution of y_t cannot depend on y_{t-1}.

Hopefully, the pattern is clear now. If we have the model

$$y_t = \beta_0 + \beta_1 z_t + \beta_2 y_{t-1} + \beta_3 z_{t-1} + u_t,$$

the homoskedasticity assumption is

$$\text{Var}(u_t|z_t,y_{t-1},z_{t-1}) = \text{Var}(y_t|z_t,y_{t-1},z_{t-1}) = \sigma^2,$$

so that the variance of u_t cannot depend on z_t, y_{t-1}, or z_{t-1} (or some other function of time). Generally, whatever explanatory variables appear in the model, we must assume that the variance of y_t given these explanatory variables is constant. If the model contains lagged y or lagged explanatory variables, then we are explicitly ruling out dynamic forms of heteroskedasticity (something we study in Chapter 12). But, in a static model, we are only concerned with $\text{Var}(y_t|z_t)$. In equation (11.37), no direct restrictions are placed on, say, $\text{Var}(y_t|y_{t-1})$.

SUMMARY

In this chapter, we have argued that OLS can be justified using asymptotic analysis, provided certain conditions are met. Ideally, the time series processes are stationary and weakly dependent, although stationarity is not crucial. Weak dependence is necessary for applying the standard large sample results, particularly the central limit theorem.

Processes with deterministic trends that are weakly dependent can be used directly in regression analysis, provided time trends are included in the model (as in Section 10.5). A similar statement holds for processes with seasonality.

When the time series are highly persistent (they have unit roots), we must exercise extreme caution in using them directly in regression models (unless we are convinced the CLM assumptions from Chapter 10 hold). An alternative to using the levels is to use the first differences of the variables. For most highly persistent economic time series, the first difference is weakly dependent. Using first differences changes the nature of the model, but this method is often as informative as a model in levels. When data are highly persistent, we usually have more faith in first-difference results. In Chapter 18, we will cover some recent, more advanced methods for using I(1) variables in multiple regression analysis.

When models have complete dynamics in the sense that no further lags of any variable are needed in the equation, we have seen that the errors will be serially uncorrelated. This is useful because certain models, such as autoregressive models, are assumed to have complete dynamics. In static and distributed lag models, the dynamically complete assumption is often false, which generally means the errors will be serially correlated. We will see how to address this problem in Chapter 12.

KEY TERMS

Asymptotically Uncorrelated
Autoregressive Process of Order One [AR(1)]
Covariance Stationary
Dynamically Complete Model
First Difference
Highly Persistent
Integrated of Order One [I(1)]
Integrated of Order Zero [I(0)]
Moving Average Process of Order One [MA(1)]
Nonstationary Process
Random Walk
Random Walk with Drift
Serially Uncorrelated
Stable AR(1) Process
Stationary Process
Strongly Dependent
Trend-Stationary Process
Unit Root Process
Weakly Dependent

PROBLEMS

11.1 Let $\{x_t: t = 1,2, \ldots\}$ be a covariance stationary process and define $\gamma_h = \text{Cov}(x_t,x_{t+h})$ for $h \geq 0$. [Therefore, $\gamma_0 = \text{Var}(x_t)$.] Show that $\text{Corr}(x_t,x_{t+h}) = \gamma_h/\gamma_0$.

11.2 Let $\{e_t: t = -1,0,1, \ldots\}$ be a sequence of independent, identically distributed random variables with mean zero and variance one. Define a stochastic process by

$$x_t = e_t - (1/2)e_{t-1} + (1/2)e_{t-2}, \ t = 1,2, \ldots.$$

(i) Find $\text{E}(x_t)$ and $\text{Var}(x_t)$. Do either of these depend on t?
(ii) Show that $\text{Corr}(x_t,x_{t+1}) = -1/2$ and $\text{Corr}(x_t,x_{t+2}) = 1/3$. (*Hint*: It is easiest to use the formula in Problem 11.1.)
(iii) What is $\text{Corr}(x_t,x_{t+h})$ for $h > 2$?
(iv) Is $\{x_t\}$ an asymptotically uncorrelated process?

11.3 Suppose that a time series process $\{y_t\}$ is generated by $y_t = z + e_t$, for all $t = 1,2, \ldots,$ where $\{e_t\}$ is an i.i.d. sequence with mean zero and variance σ_e^2. The random variable z does not change over time; it has mean zero and variance σ_z^2. Assume that each e_t is uncorrelated with z.

(i) Find the expected value and variance of y_t. Do your answers depend on t?

(ii) Find $\text{Cov}(y_t, y_{t+h})$ for any t and h. Is $\{y_t\}$ covariance stationary?

(iii) Use parts (i) and (ii) to show that $\text{Corr}(y_t, y_{t+h}) = \sigma_z^2/(\sigma_z^2 + \sigma_e^2)$ for all t and h.

(iv) Does y_t satisfy the intuitive requirement for being asymptotically uncorrelated? Explain.

11.4 Let $\{y_t: t = 1,2, \dots\}$ follow a random walk, as in (11.20), with $y_0 = 0$. Show that $\text{Corr}(y_t, y_{t+h}) = \sqrt{t/(t + h)}$ for $t \geq 1$, $h > 0$.

11.5 For the U.S. economy, let *gprice* denote the monthly growth in the overall price level and let *gwage* be the monthly growth in hourly wages. [These are both obtained as differences of logarithms: $gprice = \Delta\log(price)$ and $gwage = \Delta\log(wage)$.] Using the monthly data in WAGEPRC.RAW, we estimate the following distributed lag model:

$$\begin{aligned}\widehat{gprice} = {} & -\underset{(.00057)}{.00093} + \underset{(.052)}{.119}\, gwage + \underset{(.039)}{.097}\, gwage_{-1} + \underset{(.039)}{.040}\, gwage_{-2} \\ & + \underset{(.039)}{.038}\, gwage_{-3} + \underset{(.039)}{.081}\, gwage_{-4} + \underset{(.039)}{.107}\, gwage_{-5} + \underset{(.039)}{.095}\, gwage_{-6} \\ & + \underset{(.039)}{.104}\, gwage_{-7} + \underset{(.039)}{.103}\, gwage_{-8} + \underset{(.039)}{.159}\, gwage_{-9} + \underset{(.039)}{.110}\, gwage_{-10} \\ & + \underset{(.039)}{.103}\, gwage_{-11} + \underset{(.052)}{.016}\, gwage_{-12} \\ & n = 273,\ R^2 = .317,\ \bar{R}^2 = .283.\end{aligned}$$

(i) Sketch the estimated lag distribution. At what lag is the effect of *gwage* on *gprice* largest? Which lag has the smallest coefficient?

(ii) For which lags are the t statistics less than two?

(iii) What is the estimated long-run propensity? Is it much different than one? Explain what the LRP tells us in this example.

(iv) What regression would you run to obtain the standard error of the LRP directly?

(v) How would you test the joint significance of six more lags of *gwage*? What would be the *df*s in the F distribution? (Be careful here; you lose six more observations.)

11.6 Let $hy6_t$ denote the three-month holding yield (in percent) from buying a six-month T-bill at time $(t - 1)$ and selling it at time t (three months hence) as a three-month T-bill. Let $hy3_{t-1}$ be the three-month holding yield from buying a three-month T-bill at time $(t - 1)$. At time $(t - 1)$, $hy3_{t-1}$ is known, whereas $hy6_t$ is unknown because $p3_t$ (the price of three-month T-bills) is unknown at time $(t - 1)$. The *expectations hypothesis* (EH) says that these two different three-month investments should be the same, on average. Mathematically, we can write this as a conditional expectation:

$$\text{E}(hy6_t | I_{t-1}) = hy3_{t-1},$$

where I_{t-1} denotes all observable information up through time $t - 1$. This suggests estimating the model

$$hy6_t = \beta_0 + \beta_1 hy3_{t-1} + u_t,$$

and testing H_0: $\beta_1 = 1$. (We can also test H_0: $\beta_0 = 0$, but we often allow for a *term premium* for buying assets with different maturities, so that $\beta_0 \neq 0$.)

(i) Estimating the previous equation by OLS using the data in INTQRT.RAW (spaced every three months) gives

$$\begin{aligned} \widehat{hy6}_t &= -.058 + 1.104\ hy3_{t-1} \\ &\quad\ (.070) \quad (0.039) \\ n &= 123,\ R^2 = .866. \end{aligned}$$

Do you reject H_0: $\beta_1 = 1$ against H_0: $\beta_1 \neq 1$ at the 1% significance level? Does the estimate seem practically different from one?

(ii) Another implication of the EH is that no other variables dated as $(t - 1)$ or earlier should help explain $hy6_t$, once $hy3_{t-1}$ has been controlled for. Including one lag of the *spread* between six-month and three-month, T-bill rates gives

$$\begin{aligned} \widehat{hy6}_t &= -.123 + 1.053\ hy3_{t-1} + .480\ (r6_{t-1} - r3_{t-1}) \\ &\quad\ (.067) \quad (0.039) \qquad\qquad (.109) \\ n &= 123,\ R^2 = .885. \end{aligned}$$

Now is the coefficient on $hy3_{t-1}$ statistically different from one? Is the lagged spread term significant? According to this equation, if, at time $(t - 1)$, $r6$ is above $r3$, should you invest in six-month or three-month, T-bills?

(iii) The sample correlation between $hy3_t$ and $hy3_{t-1}$ is .914. Why might this raise some concerns with the previous analysis?

(iv) How would you test for seasonality in the equation estimated in part (ii)?

11.7 A *partial adjustment model* is

$$\begin{aligned} y_t^* &= \gamma_0 + \gamma_1 x_t + e_t \\ y_t - y_{t-1} &= \lambda(y_t^* - y_{t-1}) + a_t, \end{aligned}$$

where y_t^* is the desired or optimal level of y, and y_t is the actual (observed) level. For example, y_t^* is the desired growth in firm inventories, and x_t is growth in firm sales. The parameter γ_1 measures the effect of x_t on y_t^*. The second equation describes how the actual y adjusts depending on the relationship between the desired y in time t and the actual y in time $(t - 1)$. The parameter λ measures the speed of adjustment and satisfies $0 < \lambda < 1$.

(i) Plug the first equation for y_t^* into the second equation and show that we can write

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 x_t + u_t.$$

In particular, find the β_j in terms of the γ_j and λ and find u_t in terms of e_t and a_t. Therefore, the partial adjustment model leads to a model with a lagged dependent variable and a contemporaneous x.

(ii) If $\mathrm{E}(e_t|x_t,y_{t-1},x_{t-1},\ldots) = \mathrm{E}(a_t|x_t,y_{t-1},x_{t-1},\ldots) = 0$ and all series are weakly dependent, how would you estimate the β_j?

(iii) If $\hat{\beta}_1 = .7$ and $\hat{\beta}_2 = .2$, what are the estimates of γ_1 and λ?

COMPUTER EXERCISES

11.8 Use the data in HSEINV.RAW for this exercise.

(i) Find the first order autocorrelation in log(*invpc*). Now find the autocorrelation *after* linearly detrending log(*invpc*). Do the same for log(*price*). Which of the two series may have a unit root?

(ii) Based on your findings in part (i), estimate the equation

$$\log(invpc_t) = \beta_0 + \beta_1 \Delta\log(price_t) + \beta_2 t + u_t$$

and report the results in standard form. Interpret the coefficient $\hat{\beta}_1$ and determine whether it is statistically significant.

(iii) Linearly detrend $\log(invpc_t)$ and use the detrended version as the dependent variable in the regression from part (ii) (see Section 10.5). What happens to R^2?

(iv) Now use $\Delta\log(invpc_t)$ as the dependent variable. How do your results change from part (ii)? Is the time trend still significant? Why or why not?

11.9 In Example 11.7, define the growth in hourly wage and output per hour as the change in the natural log: $ghrwage = \Delta\log(hrwage)$ and $goutphr = \Delta\log(outphr)$. Consider a simple extension of the model estimated in (11.29):

$$ghrwage_t = \beta_0 + \beta_1 goutphr_t + \beta_2 goutphr_{t-1} + u_t.$$

This allows an increase in productivity growth to have both a current and lagged effect on wage growth.

(i) Estimate the equation using the data in EARNS.RAW and report the results in standard form. Is the lagged value of *goutphr* statistically significant?

(ii) If $\beta_1 + \beta_2 = 1$, a permanent increase in productivity growth is fully passed on in higher wage growth after one year. Test $\mathrm{H}_0\colon \beta_1 + \beta_2 = 1$ against the two-sided alternative. Remember, the easiest way to do this is to write the equation so that $\theta = \beta_1 + \beta_2$ appears directly in the model, as in Example 10.4 from Chapter 10.

(iii) Does $goutphr_{t-2}$ need to be in the model? Explain.

11.10 (i) In Example 11.4, it may be that the expected value of the return at time t, given past returns, is a quadratic function of $return_{t-1}$. To check this possibility, use the data in NYSE.RAW to estimate

$$return_t = \beta_0 + \beta_1 return_{t-1} + \beta_2 return_{t-1}^2 + u_t;$$

report the results in standard form.

(ii) State and test the null hypothesis that $E(return_t|return_{t-1})$ does not depend on $return_{t-1}$. (*Hint*: There are two restrictions to test here.) What do you conclude?

(iii) Drop $return^2_{t-1}$ from the model, but add the interaction term $return_{t-1} \cdot return_{t-2}$. Now, test the efficient markets hypothesis.

(iv) What do you conclude about predicting weekly stock returns based on past stock returns?

11.11 Use the data in PHILLIPS.RAW for this exercise.

(i) In Example 11.5, we assumed that the natural rate of unemployment is constant. An alternative form of the expectations augmented Phillips curve allows the natural rate of unemployment to depend on past levels of unemployment. In the simplest case, the natural rate at time t equals $unem_{t-1}$. If we assume adaptive expectations, we obtain a Phillips curve where inflation and unemployment are in first differences:

$$\Delta inf = \beta_0 + \beta_1 \Delta unem + u.$$

Estimate this model, report the results in the usual form, and discuss the sign, size, and statistical significance of $\hat{\beta}_1$.

(ii) Which model fits the data better, (11.19) or the model from part (i)? Explain.

11.12 (i) Add a linear time trend to equation (11.27). Is a time trend necessary in the first-difference equation?

(ii) Drop the time trend and add the variables *ww2* and *pill* to (11.27) (do not difference these dummy variables). Are these variables jointly significant at the 5% level?

(iii) Using the model from part (ii), estimate the LRP and obtain its standard error. Compare this to (10.19), where *gfr* and *pe* appeared in levels rather than in first differences.

11.13 Let $inven_t$ be the real value inventories in the United States during year t, let GDP_t denote real gross domestic product, and let $r3_t$ denote the (ex post) real interest rate on three-month T-bills. The ex post real interest rate is (approximately) $r3_t = i3_t - inf_t$, where $i3_t$ is the rate on three-month T-bills and inf_t is the annual inflation rate [see Mankiw (1994, Section 6.4)]. The change in inventories, $\Delta inven_t$, is the *inventory investment* for the year. The *accelerator model* of inventory investment is

$$\Delta inven_t = \beta_0 + \beta_1 \Delta GDP_t + u_t,$$

where $\beta_1 > 0$. [See, for example, Mankiw (1994), Chapter 17.]

(i) Use the data in INVEN.RAW to estimate the accelerator model. Report the results in the usual form and interpret the equation. Is $\hat{\beta}_1$ statistically greater than zero?

(ii) If the real interest rate rises, then the opportunity cost of holding inventories rises, and so an increase in the real interest rate should decrease inventories. Add the real interest rate to the accelerator model and discuss the results. Does the level of the real interest rate work better than the first difference, $\Delta r3_t$?

11.14 Use CONSUMP.RAW for this exercise. One version of the *permanent income hypothesis* (PIH) of consumption is that the *growth* in consumption is unpredictable. [Another version is that the change in consumption itself is unpredictable; see Mankiw (1994, Chapter 15) for discussion of the PIH.] Let $gc_t = \log(c_t) - \log(c_{t-1})$ be the growth in real per capita consumption (of nondurables and services). Then the PIH implies that $\mathrm{E}(gc_t|I_{t-1}) = \mathrm{E}(gc_t)$, where I_{t-1} denotes information known at time $(t-1)$; in this case, t denotes a year.

(i) Test the PIH by estimating $gc_t = \beta_0 + \beta_1 gc_{t-1} + u_t$. Clearly state the null and alternative hypotheses. What do you conclude?

(ii) To the regression in part (i), add gy_{t-1} and $i3_{t-1}$, where gy_t is the growth in real per capita disposable income and $i3_t$ is the interest rate on three-month T-bills; note that each must be lagged in the regression. Are these two additional variables jointly significant?

11.15 Use the data in PHILLIPS.RAW for this exercise.

(i) Estimate an AR(1) model for the unemployment rate. Use this equation to predict the unemployment rate for 1997. Compare this with the actual unemployment rate for 1997. (You can find this information in a recent *Economic Report of the President*.)

(ii) Add a lag of inflation to the AR(1) model from part (i). Is inf_{t-1} statistically significant?

(iii) Use the equation from part (ii) to predict the unemployment rate for 1997. Is the result better or worse than in the model from part (i)?

(iv) Use the method from Section 6.4 to construct a 95% prediction interval for the 1997 unemployment rate. Is the 1997 unemployment rate in the interval?

Chapter Twelve

Serial Correlation and Heteroskedasticity in Time Series Regressions

In this chapter, we discuss the critical problem of serial correlation in the error terms of a multiple regression model. We saw in Chapter 11 that when, in an appropriate sense, the dynamics of a model have been completely specified, the errors will not be serially correlated. Thus, testing for serial correlation can be used to detect dynamic misspecification. Furthermore, static and finite distributed lag models often have serially correlated errors even if there is no underlying misspecification of the model. Therefore, it is important to know the consequences and remedies for serial correlation for these useful classes of models.

In Section 12.1, we present the properties of OLS when the errors contain serial correlation. In Section 12.2, we demonstrate how to test for serial correlation. We cover tests that apply to models with strictly exogenous regressors and tests that are asymptotically valid with general regressors, including lagged dependent variables. Section 12.3 explains how to correct for serial correlation under the assumption of strictly exogenous explanatory variables, while Section 12.4 shows how using differenced data often eliminates serial correlation in the errors. Section 12.5 covers more recent advances on how to adjust the usual OLS standard errors and test statistics in the presence of very general serial correlation.

In Chapter 8, we discussed testing and correcting for heteroskedasticity in cross-sectional applications. In Section 12.6, we show how the methods used in the cross-sectional case can be extended to the time series case. The mechanics are essentially the same, but there are a few subtleties associated with the temporal correlation in time series observations that must be addressed. In addition, we briefly touch on the consequences of dynamic forms of heteroskedasticity.

12.1 PROPERTIES OF OLS WITH SERIALLY CORRELATED ERRORS

Unbiasedness and Consistency

In Chapter 10, we proved unbiasedness of the OLS estimator under the first three Gauss-Markov assumptions for time series regressions (TS.1 through TS.3). In particular, Theorem 10.1 assumed nothing about serial correlation in the errors. It follows that, as long as the explanatory variables are strictly exogenous, the $\hat{\beta}_j$ are unbiased, regardless of the degree of serial correlation in the errors. This is analogous to the observation that heteroskedasticity in the errors does not cause bias in the $\hat{\beta}_j$.

In Chapter 11, we relaxed the strict exogeneity assumption to $\mathrm{E}(u_t|\boldsymbol{x}_t) = 0$ and showed that, when the data are weakly dependent, the $\hat{\beta}_j$ are still consistent (although not necessarily unbiased). This result did not hinge on any assumption about serial correlation in the errors.

Efficiency and Inference

Since the Gauss-Markov theorem (Theorem 10.4) requires both homoskedasticity and serially uncorrelated errors, OLS is no longer BLUE in the presence of serial correlation. Even more importantly, the usual OLS standard errors and test statistics are not valid, even asymptotically. We can see this by computing the variance of the OLS estimator under the first four Gauss-Markov assumptions and the AR(1) model for the error terms. More precisely, we assume that

$$u_t = \rho u_{t-1} + e_t,\ t = 1,2,\ldots,n \tag{12.1}$$

$$|\rho| < 1, \tag{12.2}$$

where the e_t are uncorrelated random variables with mean zero and variance σ_e^2; recall from Chapter 11 that assumption (12.2) is the stability condition.

We consider the variance of the OLS slope estimator in the simple regression model

$$y_t = \beta_0 + \beta_1 x_t + u_t,$$

and, just to simplify the formula, we assume that the sample average of the x_t is zero ($\bar{x} = 0$). Then the OLS estimator $\hat{\beta}_1$ of β_1 can be written as

$$\hat{\beta}_1 = \beta_1 + \mathrm{SST}_x^{-1} \sum_{t=1}^{n} x_t u_t, \tag{12.3}$$

where $\mathrm{SST}_x = \sum_{t=1}^{n} x_t^2$. Now, in computing the variance of $\hat{\beta}_1$ (conditional on $\boldsymbol{X}$), we must account for the serial correlation in the u_t:

$$\begin{aligned}\mathrm{Var}(\hat{\beta}_1) &= \mathrm{SST}_x^{-2}\mathrm{Var}\left(\sum_{t=1}^{n} x_t u_t\right) = \mathrm{SST}_x^{-2}\left(\sum_{t=1}^{n} x_t^2 \mathrm{Var}(u_t)\right.\\ &\quad \left. + 2\sum_{t=1}^{n-1}\sum_{j=1}^{n-t} x_t x_{t+j}\mathrm{E}(u_t u_{t+j})\right)\\ &= \sigma^2/\mathrm{SST}_x + 2(\sigma^2/\mathrm{SST}_x^2)\sum_{t=1}^{n-1}\sum_{j=1}^{n-t}\rho^j x_t x_{t+j},\end{aligned} \tag{12.4}$$

where $\sigma^2 = \mathrm{Var}(u_t)$ and we have used the fact that $\mathrm{E}(u_t u_{t+j}) = \mathrm{Cov}(u_t, u_{t+j}) = \rho^j \sigma^2$ [see equation (11.4)]. The first term in equation (12.4), σ^2/SST_x, is the variance of $\hat{\beta}_1$ when $\rho = 0$, which is the familiar OLS variance under the Gauss-Markov assumptions. If we ignore the serial correlation and estimate the variance in the usual way, the variance estimator will usually be biased when $\rho \neq 0$ because it ignores the second term in (12.4). As we will see through later examples, $\rho > 0$ is most common, in which case, $\rho^j > 0$ for all j. Further, the independent variables in regression models are often positively correlated over time, so that $x_t x_{t+j}$ is positive for most pairs t and $t + j$. Therefore, in most economic applications, the term $\sum_{t=1}^{n-1} \sum_{j=1}^{n-t} \rho^j x_t x_{t+j}$ is positive, and so the usual OLS variance formula σ^2/SST_x *underestimates* the true variance of the OLS estimator. If ρ is large or x_t has a high degree of positive serial correlation—a common case—the bias in the usual OLS variance estimator can be substantial. We will tend to think the OLS slope estimator is more precise than it actually is.

When $\rho < 0$, ρ^j is negative when j is odd and positive when j is even, and so it is difficult to determine the sign of $\sum_{t=1}^{n-1} \sum_{j=1}^{n-t} \rho^j x_t x_{t+j}$. In fact, it is possible that the usual OLS variance formula actually *overstates* the true variance of $\hat{\beta}_1$. In either case, the usual variance estimator will be biased for $\mathrm{Var}(\hat{\beta}_1)$ in the presence of serial correlation.

Because the standard error of $\hat{\beta}_1$ is an estimate of the standard deviation of $\hat{\beta}_1$, using the usual OLS standard error in the presence of serial correlation is invalid. Therefore, t statistics are no longer valid for testing single hypotheses. Since a smaller standard error means a larger t statistic, the usual t statistics will often be too large when $\rho > 0$. The usual F and LM statistics for testing multiple hypotheses are also invalid.

QUESTION 12.1

Suppose that, rather than the AR(1) model, u_t follows the MA(1) model $u_t = e_t + \alpha e_{t-1}$. Find $\mathrm{Var}(\hat{\beta}_1)$ and show that it is different from the usual formula if $\alpha \neq 0$.

Serial Correlation in the Presence of Lagged Dependent Variables

Beginners in econometrics are often warned of the dangers of serially correlated errors in the presence of lagged dependent variables. Almost every textbook on econometrics contains some form of the statement "OLS is inconsistent in the presence of lagged dependent variables and serially correlated errors." Unfortunately, as a general assertion, this statement is false. There is a version of the statement that is correct, but it is important to be very precise.

To illustrate, suppose that the expected value of y_t, given y_{t-1}, is linear:

$$\mathrm{E}(y_t|y_{t-1}) = \beta_0 + \beta_1 y_{t-1}, \tag{12.5}$$

where we assume stability, $|\beta_1| < 1$. We know we can always write this with an error term as

$$y_t = \beta_0 + \beta_1 y_{t-1} + u_t, \tag{12.6}$$

$$\mathrm{E}(u_t|y_{t-1}) = 0. \tag{12.7}$$

By construction, this model satisfies the key Assumption TS.3′ for consistency of OLS, and therefore the OLS estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ are consistent. It is important to see that, without further assumptions, the errors $\{u_t\}$ *can* be serially correlated. Condition (12.7) ensures that u_t is uncorrelated with y_{t-1}, but u_t and y_{t-2} could be correlated. Then, since $u_{t-1} = y_{t-1} - \beta_0 - \beta_1 y_{t-2}$, the covariance between u_t and u_{t-1} is $-\beta_1 \text{Cov}(u_t, y_{t-2})$, which is not necessarily zero. Thus, the errors exhibit serial correlation and the model contains a lagged dependent variable, but OLS consistently estimates β_0 and β_1 because these are the parameters in the conditional expectation (12.5). The serial correlation in the errors will cause the usual OLS statistics to be invalid for testing purposes, but it will not affect consistency.

So when is OLS inconsistent if the errors are serially correlated and the regressors contain a lagged dependent variable? This happens when we write the model in error form, exactly as in (12.6), but then we *assume* that $\{u_t\}$ follows a stable AR(1) model as in (12.1) and (12.2), where

$$\text{E}(e_t|u_{t-1}, u_{t-2}, \ldots) = \text{E}(e_t|y_{t-1}, y_{t-2}, \ldots) = 0. \tag{12.8}$$

Since e_t is uncorrelated with y_{t-1} by assumption, $\text{Cov}(y_{t-1}, u_t) = \rho\text{Cov}(y_{t-1}, u_{t-1})$, which is not zero unless $\rho = 0$. This causes the OLS estimators of β_0 and β_1 from the regression of y_t on y_{t-1} to be inconsistent.

We now see that OLS estimation of (12.6), when the errors u_t also follow an AR(1) model, leads to inconsistent estimators. However, the correctness of this statement makes it no less wrongheaded. We have to ask: What would be the point in estimating the parameters in (12.6) when the errors follow an AR(1) model? It is difficult to think of cases where this would be interesting. At least in (12.5) the parameters tell us the expected value of y_t given y_{t-1}. When we combine (12.6) and (12.1), we see that y_t really follows a second order autoregressive model, or AR(2) model. To see this, write $u_{t-1} = y_{t-1} - \beta_0 - \beta_1 y_{t-2}$ and plug this into $u_t = \rho u_{t-1} + e_t$. Then, (12.6) can be rewritten as

$$\begin{aligned} y_t &= \beta_0 + \beta_1 y_{t-1} + \rho(y_{t-1} - \beta_0 - \beta_1 y_{t-2}) + e_t \\ &= \beta_0(1 - \rho) + (\beta_1 + \rho)y_{t-1} - \rho\beta_1 y_{t-2} + e_t \\ &= \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + e_t, \end{aligned}$$

where $\alpha_0 = \beta_0(1 - \rho)$, $\alpha_1 = \beta_1 + \rho$, and $\alpha_2 = -\rho\beta_1$. Given (12.8), it follows that

$$\text{E}(y_t|y_{t-1}, y_{t-2}, \ldots) = \text{E}(y_t|y_{t-1}, y_{t-2}) = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 y_{t-2}. \tag{12.9}$$

This means that the expected value of y_t, given all past y, depends on *two* lags of y. It is equation (12.9) that we would be interested in using for any practical purpose, including forecasting, as we will see in Chapter 18. We are especially interested in the parameters α_j. Under the appropriate stability conditions for an AR(2) model—we will cover these in Section 12.3—OLS estimation of (12.9) produces consistent and asymptotically normal estimators of the α_j.

The bottom line is that you need a good reason for having both a lagged dependent variable in a model and a particular model of serial correlation in the errors. Often serial correlation in the errors of a dynamic model simply indicates that the dynamic regression function has not been completely specified: in the previous example, we should add y_{t-2} to the equation.

In Chapter 18, we will see examples of models with lagged dependent variables where the errors are serially correlated and are also correlated with y_{t-1}. But even in these cases, the errors do not follow an autoregressive process.

12.2 TESTING FOR SERIAL CORRELATION

In this section, we discuss several methods of testing for serial correlation in the error terms in the multiple linear regression model

$$y_t = \beta_0 + \beta_1 x_{t1} + \dots + \beta_k x_{tk} + u_t.$$

We first consider the case when the regressors are strictly exogenous. Recall that this requires the error, u_t, to be uncorrelated with the regressors in all time periods (see Section 10.3), and so, among other things, it rules out models with lagged dependent variables.

A *t* test for AR(1) Serial Correlation with Strictly Exogenous Regressors

While there are numerous ways in which the error terms in a multiple regression model can be serially correlated, the most popular model—and the simplest to work with—is the AR(1) model in equations (12.1) and (12.2). In the previous section, we explained the implications of performing OLS when the errors are serially correlated in general, and we derived the variance of the OLS slope estimator in a simple regression model with AR(1) errors. We now show how to test for the presence of AR(1) serial correlation. The null hypothesis is that there is *no* serial correlation. Therefore, just as with tests for heteroskedasticity, we assume the best and require the data to provide reasonably strong evidence that the ideal assumption of no serial correlation is violated.

We first derive a large sample test, under the assumption that the explanatory variables are strictly exogenous: the expected value of u_t, given the entire history of independent variables, is zero. In addition, in (12.1), we must assume that

$$\mathrm{E}(e_t | u_{t-1}, u_{t-2}, \dots) = 0 \quad \textbf{(12.10)}$$

and

$$\mathrm{Var}(e_t | u_{t-1}) = \mathrm{Var}(e_t) = \sigma_e^2. \quad \textbf{(12.11)}$$

These are standard assumptions in the AR(1) model (which follow when $\{e_t\}$ is an i.i.d. sequence), and they allow us to apply the large sample results from Chapter 11 for dynamic regression.

As with testing for heteroskedasticity, the null hypothesis is that the appropriate Gauss-Markov assumption is true. In the AR(1) model, the null hypothesis that the errors are serially uncorrelated is

$$\mathrm{H}_0\text{: } \rho = 0. \quad \textbf{(12.12)}$$

How can we test this hypothesis? If the u_t were observed, then, under (12.10) and (12.11), we could immediately apply the asymptotic normality results from Theorem 11.2 to the dynamic regression model

$$u_t = \rho u_{t-1} + e_t, \ t = 2, \ldots, n. \quad \textbf{(12.13)}$$

(Under the null hypothesis $\rho = 0$, $\{u_t\}$ is clearly weakly dependent.) In other words, we could estimate ρ from the regression of u_t on u_{t-1}, for all $t = 2, \ldots, n$, without an intercept, and use the usual t statistic for $\hat{\rho}$. This does not work because the errors u_t are not observed. Nevertheless, just as with testing for heteroskedasticity, we can replace u_t with the corresponding OLS residual, $\hat{u}_t$. Since $\hat{u}_t$ depends on the OLS estimators $\hat{\beta}_0$, $\hat{\beta}_1, \ldots, \hat{\beta}_k$, it is not obvious that using $\hat{u}_t$ for u_t in the regression has no effect on the distribution of the t statistic. Fortunately, it turns out that, because of the strict exogeneity assumption, the large sample distribution of the t statistic is not affected by using the OLS residuals in place of the errors. A proof is well-beyond the scope of this text, but it follows from the work of Wooldridge (1991b).

We can summarize the asymptotic test for AR(1) serial correlation very simply:

TESTING FOR AR(1) SERIAL CORRELATION WITH STRICTLY EXOGENOUS REGRESSORS:

(i) Run the OLS regression of y_t on $x_{t1}, \ldots, x_{tk}$ and obtain the OLS residuals, $\hat{u}_t$, for all $t = 1,2, \ldots, n$.

(ii) Run the regression of

$$\hat{u}_t \text{ on } \hat{u}_{t-1}, \text{ for all } t = 2, \ldots, n, \quad \textbf{(12.14)}$$

obtaining the coefficient $\hat{\rho}$ on $\hat{u}_{t-1}$ and its t statistic, $t_{\hat{\rho}}$. (This regression may or may not contain an intercept; the t statistic for $\hat{\rho}$ will be slightly affected, but it is asymptotically valid either way.)

(iii) Use $t_{\hat{\rho}}$ to test H_0: $\rho = 0$ against H_1: $\rho \neq 0$ in the usual way. (Actually, since $\rho > 0$ is often expected a priori, the alternative can be H_0: $\rho > 0$.) Typically, we conclude that serial correlation is a problem to be dealt with only if H_0 is rejected at the 5% level. As always, it is best to report the p-value for the test.

In deciding whether serial correlation needs to be addressed, we should remember the difference between practical and statistical significance. With a large sample size, it is possible to find serial correlation even though $\hat{\rho}$ is practically small; when $\hat{\rho}$ is close to zero, the usual OLS inference procedures will not be far off [see equation (12.4)]. Such outcomes are somewhat rare in time series applications because time series data sets are usually small.

EXAMPLE 12.1
[Testing for AR(1) Serial Correlation in the Phillips Curve]

In Chapter 10, we estimated a static Phillips curve that explained the inflation-unemployment tradeoff in the United States (see Example 10.1). In Chapter 11, we studied a particular expectations augmented Phillips curve, where we assumed adaptive expectations (see Example 11.5). We now test the error term in each equation for serial correlation. Since the expectations augmented curve uses $\Delta inf_t = inf_t - inf_{t-1}$ as the dependent variable, we have one fewer observation.

For the static Phillips curve, the regression in (12.14) yields $\hat{\rho} = .573$, $t = 4.93$, and p-value = .000 (with 48 observations). This is very strong evidence of positive, first order serial correlation. One consequence of this is that the standard errors and t statistics from Chapter 10 are not valid. By contrast, the test for AR(1) serial correlation in the expectations augmented curve gives $\hat{\rho} = -.036$, $t = -.297$, and p-value = .775 (with 47 observations): there is no evidence of AR(1) serial correlation in the expectations augmented Phillips curve.

Although the test from (12.14) is derived from the AR(1) model, the test can detect other kinds of serial correlation. Remember, $\hat{\rho}$ is a consistent estimator of the correlation between u_t and u_{t-1}. Any serial correlation that causes adjacent errors to be correlated can be picked up by this test. On the other hand, it does not detect serial correlation where adjacent errors are uncorrelated, $\text{Corr}(u_t,u_{t-1}) = 0$. (For example, u_t and u_{t-2} could be correlated.)

In using the usual t statistic from (12.14), we must assume that the errors in (12.13) satisfy the appropriate homoskedasticity assumption, (12.11). In fact, it is easy to make the test robust to heteroskedasticity in e_t: we simply use the usual, heteroskedasticity-robust t statistic from Chapter 8. For the static Phillips curve in Example 12.1, the heteroskedasticity-robust t statistic is 4.03, which is smaller than the nonrobust t statistic but still very significant. In Section 12.6, we further discuss heteroskedasticity in time series regressions, including its dynamic forms.

QUESTION 12.2

How would you use regression (12.14) to construct an approximate 95% confidence interval for ρ?

The Durbin-Watson Test Under Classical Assumptions

Another test for AR(1) serial correlation is the Durbin-Watson test. The **Durbin-Watson *(DW)* statistic** is also based on the OLS residuals:

$$DW = \frac{\sum_{t=2}^{n}(\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^{n}\hat{u}_t^2}. \tag{12.15}$$

Simple algebra shows that DW and $\hat{\rho}$ from (12.14) are closely linked:

$$DW \approx 2(1 - \hat{\rho}). \tag{12.16}$$

One reason this relationship is not exact is that $\hat{\rho}$ has $\sum_{t=2}^{n}\hat{u}_{t-1}^2$ in its denominator, while the DW statistic has the sum of squares of all OLS residuals in its denominator. Even with moderate sample sizes, the approximation in (12.16) is often pretty close. Therefore, tests based on DW and the t test based on $\hat{\rho}$ are conceptually the same.

Durbin and Watson (1950) derive the distribution of *DW* (conditional on $\mathbf{X}$), something that requires the full set of classical linear model assumptions, including normality of the error terms. Unfortunately, this distribution depends on the values of the independent variables. (It also depends on the sample size, the number of regressors, and whether the regression contains an intercept.) While some econometrics packages tabulate critical values and *p*-values for *DW*, many do not. In any case, they depend on the full set of CLM assumptions.

Several econometrics texts report upper and lower bounds for the critical values that depend on the desired significance level, the alternative hypothesis, the number of observations, and the number of regressors. (We assume that an intercept is included in the model.) Usually, the *DW* test is computed for the alternative

$$\mathrm{H}_1\colon \rho > 0. \tag{12.17}$$

From the approximation in (12.16), $\hat{\rho} \approx 0$ implies that $DW \approx 2$, and $\hat{\rho} > 0$ implies that $DW < 2$. Thus, to reject the null hypothesis (12.12) in favor of (12.17), we are looking for a value of *DW* that is significantly less than two. Unfortunately, because of the problems in obtaining the null distribution of *DW*, we must compare *DW* with two sets of critical values. These are usually labelled as d_U (for *upper*) and d_L (for *lower*). If $DW < d_L$, then we reject H_0 in favor of (12.17); if $DW > d_U$, we fail to reject H_0. If $d_L \leq DW \leq d_U$, the test is inconclusive.

As an example, if we choose a 5% significance level with $n = 45$ and $k = 4$, $d_U = 1.720$ and $d_L = 1.336$ [see Savin and White (1977)]. If $DW < 1.336$, we reject the null of no serial correlation at the 5% level; if $DW > 1.72$, we fail to reject H_0; if $1.336 \leq DW \leq 1.72$, the test is inconclusive.

In Example 12.1, for the static Phillips curve, *DW* is computed to be $DW = .80$. We can obtain the lower 1% critical value from Savin and White (1977) for $k = 1$ and $n = 50$: $d_L = 1.32$. Therefore, we reject the null of no serial correlation against the alternative of positive serial correlation at the 1% level. (Using the previous *t* test, we can conclude that the *p*-value equals zero to three decimal places.) For the expectations augmented Phillips curve, $DW = 1.77$, which is well within the fail-to-reject region at even the 5% level ($d_U = 1.59$).

The fact that an exact sampling distribution for *DW* can be tabulated is the only advantage that *DW* has over the *t* test from (12.14). Given that the tabulated critical values are exactly valid only under the full set of CLM assumptions and that they can lead to a wide inconclusive region, the practical disadvantages of the *DW* are substantial. The *t* statistic from (12.14) is simple to compute and asymptotically valid without normally distributed errors. The *t* statistic is also valid in the presence of heteroskedasticity that depends on the x_{tj}; and it is easy to make it robust to any form of heteroskedasticity.

Testing for AR(1) Serial Correlation without Strictly Exogenous Regressors

When the explanatory variables are not strictly exogenous, so that one or more x_{tj} is correlated with u_{t-1}, neither the *t* test from regression (12.14) nor the Durbin-Watson statistic are valid, even in large samples. The leading case of nonstrictly exogenous regressors occurs when the model contains a lagged dependent variable: y_{t-1} and u_{t-1} are obviously correlated. Durbin (1970) suggested two alternatives to the *DW* statistic when the model contains a lagged dependent variable and the other regressors are nonrandom (or, more generally, strictly exogenous). The first is called *Durbin's h statistic*. This statistic has a practical drawback in that it cannot always be computed, and so we do not cover it here.

Durbin's alternative statistic is simple to compute and is valid when there are any number of non-strictly exogenous explanatory variables. The test also works if the explanatory variables happen to be strictly exogenous.

TESTING FOR SERIAL CORRELATION WITH GENERAL REGRESSORS:

(i) Run the OLS regression of y_t on $x_{t1}, \ldots, x_{tk}$ and obtain the OLS residuals, $\hat{u}_t$, for all $t = 1,2, \ldots, n$.

(ii) Run the regression of

$$u_t \text{ on } x_{t1}, x_{t2}, \ldots, x_{tk}, \hat{u}_{t-1}, \text{ for all } t = 2, \ldots, n. \tag{12.18}$$

to obtain the coefficient $\hat{\rho}$ on $\hat{u}_{t-1}$ and its t statistic, $t_{\hat{\rho}}$.

(iii) Use $t_{\hat{\rho}}$ to test H_0: $\rho = 0$ against H_1: $\rho \neq 0$ in the usual way (or use a one-sided alternative).

In equation (12.18), we regress the OLS residuals on *all* independent variables, including an intercept, and the lagged residual. The t statistic on the lagged residual is a valid test of (12.12) in the AR(1) model (12.13) (when we add $\text{Var}(u_t|\mathbf{x}_t,u_{t-1}) = \sigma^2$ under H_0). Any number of lagged dependent variables may appear among the x_{tj}, and other nonstrictly exogenous explanatory variables are allowed as well.

The inclusion of $x_{t1}, \ldots, x_{tk}$ explicitly allows for each x_{tj} to be correlated with u_{t-1}, and this ensures that $t_{\hat{\rho}}$ has an approximate t distribution in large samples. The t statistic from (12.14) ignores possible correlation between x_{tj} and u_{t-1}, so it is not valid without strictly exogenous regressors. Incidentally, because $\hat{u}_t = y_t - \hat{\beta}_0 - \hat{\beta}_1 x_{t1} - \ldots - \hat{\beta}_k x_{tk}$, it can be shown that the t statistic on $\hat{u}_{t-1}$ is the same if y_t is used in place of $\hat{u}_t$ as the dependent variable in (12.18).

The t statistic from (12.18) is easily made robust to heteroskedasticity of unknown form (in particular, when $\text{Var}(u_t|\mathbf{x}_t,u_{t-1})$ is not constant): just use the heteroskedasticity-robust t statistic on $\hat{u}_{t-1}$.

EXAMPLE 12.2
[Testing for AR(1) Serial Correlation in the Minimum Wage Equation]

In Chapter 10 (see Example 10.9), we estimated the effect of the minimum wage on the Puerto Rican employment rate. We now check whether the errors appear to contain serial correlation, using the test that does not assume strict exogeneity of the minimum wage or GNP variables. [We add the log of Puerto Rican real GNP to equation (10.38), as in Problem

10.9]. We are assuming that the underlying stochastic processes are weakly dependent, but we allow them to contain a linear time trend (by including t in the regression).

Letting $\hat{u}_t$ denote the OLS residuals, we run the regression of

$$u_t \text{ on } \log(mincov_t), \log(prgnp_t), \log(usgnp_t), t, \text{ and } \hat{u}_{t-1},$$

using the 37 available observations. The estimated coefficient on $\hat{u}_{t-1}$ is $\hat{\rho} = .481$ with $t = 2.89$ (two-sided p-value $= .007$). Therefore, there is strong evidence of AR(1) serial correlation in the errors, which means the t statistics for the $\hat{\beta}_j$ that we obtained before are not valid for inference. Remember, though, the $\hat{\beta}_j$ are still consistent if u_t is contemporaneously uncorrelated with each explanatory variable. Incidentally, if we use regression (12.14) instead, we obtain $\hat{\rho} = .417$ and $t = 2.63$, so the outcome of the test is similar in this case.

Testing for Higher Order Serial Correlation

The test from (12.18) is easily extended to higher orders of serial correlation. For example, suppose that we wish to test

$$\text{H}_0\text{: } \rho_1 = 0, \rho_2 = 0 \tag{12.19}$$

in the AR(2) model,

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + e_t.$$

This alternative model of serial correlation allows us to test for *second order serial correlation.* As always, we estimate the model by OLS and obtain the OLS residuals, $\hat{u}_t$. Then, we can run the regression of

$$\hat{u}_t \text{ on } x_{t1}, x_{t2}, \ldots, x_{tk}, \hat{u}_{t-1}, \text{ and } \hat{u}_{t-2}, \text{ for all } t = 3, \ldots, n,$$

to obtain the F test for joint significance of $\hat{u}_{t-1}$ and $\hat{u}_{t-2}$. If these two lags are jointly significant at a small enough level, say 5%, then we reject (12.19) and conclude that the errors are serially correlated.

More generally, we can test for serial correlation in the autoregressive model of order q:

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \ldots + \rho_q u_{t-q} + e_t. \tag{12.20}$$

The null hypothesis is

$$\text{H}_0\text{: } \rho_1 = 0, \rho_2 = 0, \ldots, \rho_q = 0. \tag{12.21}$$

TESTING FOR AR(q) SERIAL CORRELATION:

(i) Run the OLS regression of y_t on $x_{t1}, \ldots, x_{tk}$ and obtain the OLS residuals, $\hat{u}_t$, for all $t = 1,2, \ldots, n$.

(ii) Run the regression of

$$\hat{u}_t \text{ on } x_{t1}, x_{t2}, \ldots, x_{tk}, \hat{u}_{t-1}, \hat{u}_{t-2}, \ldots, \hat{u}_{t-q}, \text{ for all } t = (q+1), \ldots, n. \quad \textbf{(12.22)}$$

(iii) Compute the F test for joint significance of $\hat{u}_{t-1}, \hat{u}_{t-2}, \ldots, \hat{u}_{t-q}$ in (12.22). [The F statistic with y_t as the dependent variable in (12.22) can also be used, as it gives an identical answer.]

If the x_{tj} are assumed to be strictly exogenous, so that each x_{tj} is uncorrelated with u_{t-1}, $u_{t-2}, \ldots, u_{t-q}$, then the x_{tj} can be omitted from (12.22). Including the x_{tj} in the regression makes the test valid with or without the strict exogeneity assumption. The test requires the homoskedasticity assumption

$$\text{Var}(u_t|\mathbf{x}_t, u_{t-1}, \ldots, u_{t-q}) = \sigma^2. \quad \textbf{(12.23)}$$

A heteroskedasticity-robust version can be computed as described in Chapter 8.

An alternative to computing the F test is to use the Lagrange multiplier (LM) form of the statistic. (We covered the LM statistic for testing exclusion restrictions in Chapter 5 for cross-sectional analysis.) The LM statistic for testing (12.21) is simply

$$LM = (n - q)R^2_{\hat{u}}, \quad \textbf{(12.24)}$$

where $R^2_{\hat{u}}$ is just the usual R-squared from regression (12.22). Under the null hypothesis, $LM \overset{a}{\sim} \chi^2_q$. This is usually called the **Breusch-Godfrey test** for AR(q) serial correlation. The LM statistic also requires (12.23), but it can be made robust to heteroskedasticity. [For details, see Wooldridge (1991b).

EXAMPLE 12.3
[Testing for AR(3) Serial Correlation]

In the event study of the barium chloride industry (see Example 10.5), we used monthly data, so we may wish to test for higher orders of serial correlation. For illustration purposes, we test for AR(3) serial correlation in the errors underlying equation (10.22). Using regression (12.22), the F statistic for joint significance of $\hat{u}_{t-1}$, $\hat{u}_{t-2}$, and $\hat{u}_{t-3}$ is $F = 5.12$. Originally, we had $n = 131$, and we lose three observations in the auxiliary regression (12.22). Because we estimate 10 parameters in (12.22) for this example, the df in the F statistic are 3 and 118. The p-value of the F statistic is .0023, so there is strong evidence of AR(3) serial correlation.

With quarterly or monthly data that have not been seasonally adjusted, we sometimes wish to test for seasonal forms of serial correlation. For example, with quarterly data, we might postulate the autoregressive model

$$u_t = \rho_4 u_{t-4} + e_t. \quad \textbf{(12.25)}$$

From the AR(1) serial correlation tests, it is pretty clear how to proceed. When the regressors are strictly exogenous, we can use a t test on $\hat{u}_{t-4}$ in the regression of

$$\hat{u}_t \text{ on } \hat{u}_{t-4}, \text{ for all } t = 5, \ldots, n.$$

A modification of the Durbin-Watson statistic is also available [see Wallis (1972)]. When the x_{tj} are not strictly exogenous, we can use the regression in (12.18), with $\hat{u}_{t-4}$ replacing $\hat{u}_{t-1}$.

In Example 12.3, the data are monthly and are not seasonally adjusted. Therefore, it makes sense to test for correlation between u_t and u_{t-12}. A regression of $\hat{u}_t$ on $\hat{u}_{t-12}$ yields $\hat{\rho}_{12} = -.187$ and p-value $= .028$, so there is evidence of *negative* seasonal autocorrelation. (Including the regressors changes things only modestly: $\hat{\rho}_{12} = -.170$ and p-value $= .052$.) This is somewhat unusual and does not have an obvious explanation.

QUESTION 12.3

Suppose you have quarterly data and you want to test for the presence of first order or fourth order serial correlation. With strictly exogenous regressors, how would you proceed?

12.3 CORRECTING FOR SERIAL CORRELATION WITH STRICTLY EXOGENOUS REGRESSORS

If we detect serial correlation after applying one of the tests in Section 12.2, we have to do something about it. If our goal is to estimate a model with complete dynamics, we need to respecify the model. In applications where our goal is not to estimate a fully dynamic model, we need to find a way to carry out statistical inference: as we saw in Section 12.1, the usual OLS test statistics are no longer valid. In this section, we begin with the important case of AR(1) serial correlation. The traditional approach to this problem assumes fixed regressors. What are actually needed are strictly exogenous regressors. Therefore, at a minimum, we should not use these corrections when the explanatory variables include lagged dependent variables.

Obtaining the Best Linear Unbiased Estimator in the AR(1) Model

We assume the Gauss-Markov Assumptions TS.1 through TS.4, but we relax Assumption TS.5. In particular, we assume that the errors follow the AR(1) model

$$u_t = \rho u_{t-1} + e_t, \text{ for all } t = 1,2, \ldots. \quad \textbf{(12.26)}$$

Remember that Assumption TS.2 implies that u_t has a zero mean conditional on $\boldsymbol{X}$. In the following analysis, we let the conditioning on $\boldsymbol{X}$ be implied in order to simplify the notation. Thus, we write the variance of u_t as

$$\text{Var}(u_t) = \sigma_e^2/(1 - \rho^2). \quad \textbf{(12.27)}$$

For simplicity, consider the case with a single explanatory variable:

$$y_t = \beta_0 + \beta_1 x_t + u_t, \text{ for all } t = 1,2, \ldots, n.$$

Since the problem in this equation is serial correlation in the u_t, it makes sense to transform the equation to eliminate the serial correlation. For $t \geq 2$, we write

$$y_{t-1} = \beta_0 + \beta_1 x_{t-1} + u_{t-1}$$
$$y_t = \beta_0 + \beta_1 x_t + u_t.$$

Now, if we multiply this first equation by ρ and subtract it from the second equation, we get

$$y_t - \rho y_{t-1} = (1 - \rho)\beta_0 + \beta_1(x_t - \rho x_{t-1}) + e_t,\ t \geq 2,$$

where we have used the fact that $e_t = u_t - \rho u_{t-1}$. We can write this as

$$\tilde{y}_t = (1 - \rho)\beta_0 + \beta_1 \tilde{x}_t + e_t,\ t \geq 2, \quad \textbf{(12.28)}$$

where

$$\tilde{y}_t = y_t - \rho y_{t-1},\ \tilde{x}_t = x_t - \rho x_{t-1} \quad \textbf{(12.29)}$$

are called the **quasi-differenced data**. (If $\rho = 1$, these are differenced data, but remember we are assuming $|\rho| < 1$.) The error terms in (12.28) are serially uncorrelated; in fact, this equation satisfies all of the Gauss-Markov assumptions. This means that, if we knew ρ, we could estimate β_0 and β_1 by regressing $\tilde{y}_t$ on $\tilde{x}_t$, provided we divide the estimated intercept by $(1 - \rho)$.

The OLS estimators from (12.28) are not quite BLUE because they do not use the first time period. This is easily fixed by writing the equation for $t = 1$ as

$$y_1 = \beta_0 + \beta_1 x_1 + u_1. \quad \textbf{(12.30)}$$

Since each e_t is uncorrelated with u_1, we can add (12.30) to (12.28) and still have serially uncorrelated errors. However, using (12.27), $\text{Var}(u_1) = \sigma_e^2/(1 - \rho^2) > \sigma_e^2 = \text{Var}(e_t)$. [Equation (12.27) clearly does not hold when $|\rho| \geq 1$, which is why we assume the stability condition.] Thus, we must multiply (12.30) by $(1 - \rho^2)^{1/2}$ to get errors with the same variance:

$$(1 - \rho^2)^{1/2} y_1 = (1 - \rho^2)^{1/2}\beta_0 + \beta_1(1 - \rho^2)^{1/2} x_1 + (1 - \rho^2)^{1/2} u_1$$

or

$$\tilde{y}_1 = (1 - \rho^2)^{1/2}\beta_0 + \beta_1 \tilde{x}_1 + \tilde{u}_1, \quad \textbf{(12.31)}$$

where $\tilde{u}_1 = (1 - \rho^2)^{1/2} u_1$, $\tilde{y}_1 = (1 - \rho^2)^{1/2} y_1$, and so on. The error in (12.31) has variance $\text{Var}(\tilde{u}_1) = (1 - \rho^2)\text{Var}(u_1) = \sigma_e^2$, so we can use (12.31) along with (12.28) in an OLS regression. This gives the BLUE estimators of β_0 and β_1 under Assumptions TS.1 through TS.4 and the AR(1) model for u_t.This is another example of a *generalized least squares* (or GLS) estimator. We saw other GLS estimators in the context of heteroskedasticity in Chapter 8.

Adding more regressors changes very little. For $t \geq 2$, we use the equation

$$\tilde{y}_t = (1 - \rho)\beta_0 + \beta_1 \tilde{x}_{t1} + \ldots + \beta_k \tilde{x}_{tk} + e_t, \quad \textbf{(12.32)}$$

where $\tilde{x}_{tj} = x_{tj} - \rho x_{t-1,j}$. For $t = 1$, we have $\tilde{y}_1 = (1 - \rho^2)^{1/2} y_1$, $\tilde{x}_{1j} = (1 - \rho^2)^{1/2} x_{1j}$, and the intercept is $(1 - \rho^2)^{1/2}\beta_0$. For given ρ, it is fairly easy to transform the data and to carry out OLS. Unless $\rho = 0$, the GLS estimator, that is, OLS on the transformed data, will generally be different from the original OLS estimator. The GLS estimator turns out to be BLUE, and, since the errors in the transformed equation are serially uncorrelated and homoskedastic, t and F statistics from the transformed equation are valid (at least asymptotically, and exactly if the errors e_t are normally distributed).

Feasible GLS Estimation with AR(1) Errors

The problem with the GLS estimator is that ρ is rarely known in practice. However, we already know how to get a consistent estimator of ρ: we simply regress the OLS residuals on their lagged counterparts, exactly as in equation (12.14). Next, we use this estimate, $\hat{\rho}$, in place of ρ to obtain the quasi-differenced variables. We then use OLS on the equation

$$\tilde{y}_t = \beta_0 \tilde{x}_{t0} + \beta_1 \tilde{x}_{t1} + \ldots + \beta_k \tilde{x}_{tk} + error_t, \quad \textbf{(12.33)}$$

where $\tilde{x}_{t0} = (1 - \hat{\rho})$ for $t \geq 2$, and $\tilde{x}_{10} = (1 - \hat{\rho}^2)^{1/2}$. This results in the **feasible GLS (FGLS)** estimator of the β_j. The error term in (12.33) contains e_t and also the terms involving the estimation error in $\hat{\rho}$. Fortunately, the estimation error in $\hat{\rho}$ does not affect the asymptotic distribution of the FGLS estimators.

FEASIBLE GLS ESTIMATION OF THE AR(1) MODEL:

(i) Run the OLS regression of y_t on $x_{t1}, \ldots, x_{tk}$ and obtain the OLS residuals, $\hat{u}_t$, $t = 1, 2, \ldots, n$.

(ii) Run the regression in equation (12.14) and obtain $\hat{\rho}$.

(iii) Apply OLS to equation (12.33) to estimate $\beta_0, \beta_1, \ldots, \beta_k$. The usual standard errors, t statistics, and F statistics are asymptotically valid.

The cost of using $\hat{\rho}$ in place of ρ is that the feasible GLS estimator has no tractable finite sample properties. In particular, it is not unbiased, although it is consistent when the data are weakly dependent. Further, even if e_t in (12.32) is normally distributed, the t and F statistics are only approximately t and F distributed because of the estimation error in $\hat{\rho}$. This is fine for most purposes, although we must be careful with small sample sizes.

Since the FGLS estimator is not unbiased, we certainly cannot say it is BLUE. Nevertheless, it is asymptotically more efficient than the OLS estimator when the AR(1) model for serial correlation holds (and the explanatory variables are strictly exogenous). Again, this statement assumes that the time series are weakly dependent.

There are several names for FGLS estimation of the AR(1) model that come from different methods of estimating ρ and different treatment of the first observation. **Cochrane-Orcutt (CO) estimation** omits the first observation and uses $\hat{\rho}$ from (12.14), whereas **Prais-Winsten (PW) estimation** uses the first observation in the previously suggested way. Asymptotically, it makes no difference whether or not the first observation is used, but many time series samples are small, so the differences can be notable in applications.

In practice, both the Cochrane-Orcutt and Prais-Winsten methods are used in an iterative scheme. Once the FGLS estimator is found using $\hat{\rho}$ from (12.14), we can compute a new set of residuals, obtain a new estimator of ρ from (12.14), transform the data using the new estimate of ρ, and estimate (12.33) by OLS. We can repeat the whole process many times, until the estimate of ρ changes by very little from the previous iteration. Many regression packages implement an iterative procedure automatically, so there is no additional work for us. It is difficult to say whether more than one iteration helps. It seems to be helpful in some cases, but, theoretically, the large sample properties of the iterated estimator are the same as the estimator that uses only the first iteration. For details on these and other methods, see Davidson and MacKinnon (1993, Chapter 10).

EXAMPLE 12.4

(Cochrane-Orcutt Estimation in the Event Study)

We estimate the equation in Example 10.5 using iterated Cochrane-Orcutt estimation. For comparison, we also present the OLS results in Table 12.1.

The coefficients that are statistically significant in the Cochrane-Orcutt estimation do not differ by much from the OLS estimates [in particular, the coefficients on log(*chempi*), log(*rtwex*), and *afdec6*]. It is not surprising for statistically insignificant coefficients to change, perhaps markedly, across different estimation methods.

Notice how the standard errors in the second column are uniformly higher than the standard errors in column (1). This is common. The Cochrane-Orcutt standard errors account for serial correlation; the OLS standard errors do not. As we saw in Section 12.1, the OLS standard errors usually understate the actual sampling variation in the OLS estimates and should not be relied upon when significant serial correlation is present. Therefore, the effect on Chinese imports after the International Trade Commissions decision is now less statistically significant than we thought ($t_{afdec6} = -1.68$).

The Cochrane-Orcutt (CO) method reports one fewer observation than OLS; this reflects the fact that the first transformed observation is not used in the CO method. This slightly affects the degrees of freedom that are used in hypothesis tests.

Finally, an R-squared is reported for the CO estimation, which is well-below the R-squared for the OLS estimation in this case. However, these R-squareds should not be compared. For OLS, the R-squared, as usual, is based on the regression with the untransformed dependent and independent variables. For CO, the R-squared comes from the final regression of the *transformed* dependent variable on the transformed independent variables. It is not clear what this R^2 is actually measuring, nevertheless, it is traditionally reported.

Comparing OLS and FGLS

In some applications of the Cochrane-Orcutt or Prais-Winsten methods, the FGLS estimates differ in practically important ways from the OLS estimates. (This was not the

Table 12.1

Dependent Variable: log(*chnimp*)

Coefficient	OLS	Cochrane-Orcutt
log(*chempi*)	3.12 (0.48)	2.95 (0.65)
log(*gas*)	.196 (.907)	1.05 (0.99)
log(*rtwex*)	.983 (.400)	1.14 (0.51)
befile6	.060 (.261)	−.016 (.321)
affile6	−.032 (.264)	−.033 (.323)
afdec6	−.565 (.286)	−.577 (.343)
intercept	−17.70 (20.05)	−37.31 (23.22)
$\hat{\rho}$	———	.293 (.084)
Observations *R*-Squared	131 .305	130 .193

case in Example 12.4.) Typically, this has been interpreted as a verification of feasible GLS's superiority over OLS. Unfortunately, things are not so simple. To see why, consider the regression model

$$y_t = \beta_0 + \beta_1 x_t + u_t,$$

where the time series processes are stationary. Now, assuming that the law of large numbers holds, consistency of OLS for β_1 holds if

$$\text{Cov}(x_t, u_t) = 0. \quad \textbf{(12.34)}$$

Earlier, we asserted that FGLS was consistent under the strict exogeneity assumption, which is more restrictive than (12.34). In fact, it can be shown that the weakest assumption that must hold for FGLS to be consistent, *in addition to* (12.34), is that the sum of x_{t-1} and x_{t+1} is uncorrelated with u_t:

$$\text{Cov}(\{x_{t-1} + x_{t+1}\}, u_t) = 0. \tag{12.35}$$

Practically speaking, consistency of FGLS requires u_t to be uncorrelated with x_{t-1}, x_t, and x_{t+1}.

This means that OLS and FGLS might give significantly different estimates because (12.35) fails. In this case, OLS—which is still consistent under (12.34)—is preferred to FGLS (which is inconsistent). If x has a lagged effect on y, or x_{t+1} reacts to changes in u_t, FGLS can produce misleading results.

Since OLS and FGLS are different estimation procedures, we never expect them to give the same estimates. If they provide similar estimates of the β_j, then FGLS is preferred if there is evidence of serial correlation, because the estimator is more efficient and the FGLS test statistics are at least asymptotically valid. A more difficult problem arises when there are practical differences in the OLS and FGLS estimates: it is hard to determine whether such differences are statistically significant. The general method proposed by Hausman (1978) can be used, but this is beyond the scope of this text.

Consistency and asymptotic normality of OLS and FGLS rely heavily on the time series processes y_t and the x_{tj} being weakly dependent. Strange things can happen if we apply either OLS or FGLS when some processes have unit roots. We discuss this further in Chapter 18.

EXAMPLE 12.5
(Static Phillips Curve)

Table 12.2 presents OLS and iterated Cochrane-Orcutt estimates of the static Phillips curve from Example 10.1.

Table 12.2

Dependent Variable: *inf*

Coefficient	OLS	Cochrane-Orcutt
unem	.468 (.289)	−.665 (.320)
intercept	1.424 (1.719)	7.580 (2.379)
$\hat{\rho}$	———	.774 (.091)
Observations *R*-Squared	49 .053	48 .086

The coefficient of interest is on *unem*, and it differs markedly between CO and OLS. Since the CO estimate is consistent with the inflation-unemployment tradeoff, our tendency is to focus on the CO estimates. In fact, these estimates are fairly close to what is obtained by first differencing both *inf* and *unem* (see Problem 11.11), which makes sense because the quasi-differencing used in CO with $\hat{\rho} = .774$ is similar to first differencing. It may just be that *inf* and *unem* are not related in levels, but they have a negative relationship in first differences.

Correcting for Higher Order Serial Correlation

It is also possible to correct for higher orders of serial correlation. A general treatment is given in Harvey (1990). Here, we illustrate the approach for AR(2) serial correlation:

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + e_t,$$

where $\{e_t\}$ satisfies the assumptions stated for the AR(1) model. The stability condition is more complicated now. They can be shown to be [see Harvey (1990)]

$$\rho_2 > -1,\ \rho_2 - \rho_1 < 1, \text{ and } \rho_1 + \rho_2 < 1.$$

For example, the model is stable if $\rho_1 = .8$ and $\rho_2 = -.3$; the model is unstable if $\rho_1 = .7$ and $\rho_2 = .4$.

Assuming the stability conditions hold, we can obtain the transformation that eliminates the serial correlation. In the simple regression model, this is easy when $t > 2$:

$$y_t - \rho_1 y_{t-1} - \rho_2 y_{t-2} = \beta_0(1 - \rho_1 - \rho_2) + \beta_1(x_t - \rho_1 x_{t-1} - \rho_2 x_{t-2}) + e_t$$

or

$$\tilde{y}_t = \beta_0(1 - \rho_1 - \rho_2) + \beta_1 \tilde{x}_t + e_t,\ t = 3,4, \ldots, n. \tag{12.36}$$

If we know ρ_1 and ρ_2, we can easily estimate this equation by OLS after obtaining the transformed variables. Since we rarely know ρ_1 and ρ_2, we have to estimate them. As usual, we can use the OLS residuals, $\hat{u}_t$: obtain $\hat{\rho}_1$ and $\hat{\rho}_2$ from the regression of

$$\hat{u}_t \text{ on } \hat{u}_{t-1},\ \hat{u}_{t-2},\ t = 3, \ldots, n.$$

[This is the same regression used to test for AR(2) serial correlation with strictly exogenous regressors.] Then, we use $\hat{\rho}_1$ and $\hat{\rho}_2$ in place of ρ_1 and ρ_2 to obtain the transformed variables. This gives one version of the feasible GLS estimator. If we have multiple explanatory variables, then each one is transformed by $\tilde{x}_{tj} = x_{tj} - \hat{\rho}_1 x_{t-1,j} - \hat{\rho}_2 x_{t-2,j}$, when $t > 2$.

The treatment of the first two observations is a little tricky. It can be shown that the dependent variable and each independent variable (including the intercept) should be transformed by

$$\tilde{z}_1 = \{(1 + \rho_2)[(1 - \rho_2)^2 - \rho_1^2]/(1 - \rho_2)\}^{1/2} z_1$$

$$\tilde{z}_2 = (1 - \rho_2^2)^{1/2} z_2 - \{\rho_1(1 - \rho_1^2)^{1/2}/(1 - \rho_2)\} z_1,$$

where z_1 and z_2 denote either the dependent or an independent variable at $t = 1$ and $t = 2$, respectively. We will not derive these transformations. Briefly, they eliminate the serial correlation between the first two observations and make their error variances equal to σ_e^2.

Fortunately, econometrics packages geared toward time series analysis easily estimate models with general AR(q) errors; we rarely need to directly compute the transformed variables ourselves.

12.4 DIFFERENCING AND SERIAL CORRELATION

In Chapter 11, we presented differencing as a transformation for making an integrated process weakly dependent. There is another way to see the merits of differencing when dealing with highly persistent data. Suppose that we start with the simple regression model:

$$y_t = \beta_0 + \beta_1 x_t + u_t, t = 1,2, \ldots, \quad \textbf{(12.37)}$$

where u_t follows the AR(1) process (12.26). As we mentioned in Section 11.3, and as we will discuss more fully in Chapter 18, the usual OLS inference procedures can be very misleading when the variables y_t and x_t are integrated of order one, or I(1). In the extreme case where the errors $\{u_t\}$ in (12.37) follow a random walk, the equation makes no sense because, among other things, the variance of u_t grows with t. It is more logical to difference the equation:

$$\Delta y_t = \beta_1 \Delta x_t + \Delta u_t, t = 2, \ldots, n. \quad \textbf{(12.38)}$$

If u_t follows a random walk, then $e_t \equiv \Delta u_t$ has zero mean, a constant variance, and is serially uncorrelated. Thus, assuming that e_t and Δx_t are uncorrelated, we can estimate (12.38) by OLS, where we lose the first observation.

Even if u_t does not follow a random walk, but ρ is positive and large, first differencing is often a good idea: it will eliminate most of the serial correlation. Of course, (12.38) is different from (12.37), but at least we can have more faith in the OLS standard errors and t statistics in (12.38). Allowing for multiple explanatory variables does not change anything.

EXAMPLE 12.6
(Differencing the Interest Rate Equation)

In Example 10.2, we estimated an equation relating the three-month, T-bill rate to inflation and the federal deficit [see equation (10.15)]. If we regress the residuals from this equation on a single lag, we obtain $\hat{\rho} = .530$ (.123), which is statistically greater than zero. If we difference *i3*, *inf*, and *def* and then check the residuals for AR(1) serial correlation, we obtain $\hat{\rho} = .068$ (.145), and so there is no evidence of serial correlation. The differencing has apparently eliminated any serial correlation. [In addition, there is evidence that *i3* contains a unit root, and *inf* may as well, so differencing might be needed to produce I(0) variables anyway.]

QUESTION 12.4

Suppose after estimating a model by OLS that you estimate ρ from regression (12.14) and you obtain $\hat{\rho} = .92$. What would you do about this?

As we explained in Chapter 11, the decision of whether or not to difference is a tough one. But this discussion points out another benefit of differencing, which is that it removes serial correlation. We will come back to this issue in Chapter 18.

12.5 SERIAL CORRELATION-ROBUST INFERENCE AFTER OLS

In recent years, it has become more popular to estimate models by OLS but to correct the standard errors for fairly arbitrary forms of serial correlation (and heteroskedasticity). Even though we know OLS will be inefficient, there are some good reasons for taking this approach. First, the explanatory variables may not be strictly exogenous. In this case, FGLS is not even consistent, let alone efficient. Second, in most applications of FGLS, the errors are assumed to follow an AR(1) model. It may be better to compute standard errors for the OLS estimates that are robust to more general forms of serial correlation.

To get the idea, consider equation (12.4), which is the variance of the OLS slope estimator in a simple regression model with AR(1) errors. We can estimate this variance very simply by plugging in our standard estimators of ρ and σ^2. The only problem with this is that it assumes the AR(1) model holds and also homoskedasticity. It is possible to relax both of these assumptions.

A general treatment of standard errors that are both heteroskedasticity and serial correlation-robust is given in Davidson and MacKinnon (1993). Right now, we provide a simple method to compute the robust standard error of any OLS coefficient.

Our treatment here follows Wooldridge (1989). Consider the standard multiple linear regression model

$$y_t = \beta_0 + \beta_1 x_{t1} + \ldots + \beta_k x_{tk} + u_t,\ t=1,2,\ldots,n, \qquad \textbf{(12.39)}$$

which we have estimated by OLS. For concreteness, we are interested in obtaining a serial correlation-robust standard error for $\hat{\beta}_1$. This turns out to be fairly easy. Write x_{t1} as a linear function of the remaining independent variables and an error term,

$$x_{t1} = \delta_0 + \delta_2 x_{t2} + \ldots + \delta_k x_{tk} + r_t, \qquad \textbf{(12.40)}$$

where the error r_t has zero mean and is uncorrelated with $x_{t2}, x_{t3}, \ldots, x_{tk}$.

Then, it can be shown that the asymptotic variance of the OLS estimator $\hat{\beta}_1$ is

$$\text{Avar}(\hat{\beta}_1) = \left(\sum_{t=1}^{n} \text{E}(r_t^2)\right)^{-2} \text{Var}\left(\sum_{t=1}^{n} r_t u_t\right).$$

Under the no serial correlation Assumption TS.5′, $\{a_t \equiv r_t u_t\}$ is serially uncorrelated, and so either the usual OLS standard errors (under homoskedasticity) or the heteroskedasticity-robust standard errors will be valid. But if TS.5′ fails, our expression for $\text{Avar}(\hat{\beta}_1)$ must account for the correlation between a_t and a_s, when $t \neq s$. In practice, it is common to assume that, once the terms are farther apart than a few periods, the correlation is essentially zero. Remember that under weak dependence, the correlation must be approaching zero, so this is a reasonable approach.

Following the general framework of Newey and West (1987), Wooldridge (1989) shows that Avar($\hat{\beta}_1$) can be estimated as follows. Let "se($\hat{\beta}_1$)" denote the usual (but incorrect) OLS standard error and let $\hat{\sigma}$ be the usual standard error of the regression (or root mean squared error) from estimating (12.39) by OLS. Let $\hat{r}_t$ denote the residuals from the auxiliary regression of

$$x_{t1} \text{ on } x_{t2}, x_{t3}, \ldots, x_{tk} \tag{12.41}$$

(including a constant, as usual). For a chosen integer $g > 0$, define

$$\hat{v} = \sum_{t=1}^{n} \hat{a}_t^2 + 2 \sum_{h=1}^{g} [1 - h/(g+1)] \left(\sum_{t=h+1}^{n} \hat{a}_t \hat{a}_{t-h} \right), \tag{12.42}$$

where

$$\hat{a}_t = \hat{r}_t \hat{u}_t, \; t = 1,2, \ldots, n.$$

This looks somewhat complicated, but in practice it is easy to obtain. The integer g in (12.42) controls how much serial correlation we are allowing in computing the standard error. Once we have $\hat{v}$, the **serial correlation-robust standard error** of $\hat{\beta}_1$ is simply

$$\text{se}(\hat{\beta}_1) = [\text{"se}(\hat{\beta}_1)\text{"}/\hat{\sigma}]^2 \sqrt{\hat{v}}. \tag{12.43}$$

In other words, we take the usual OLS standard error of $\hat{\beta}_1$, divide it by $\hat{\sigma}$, square the result, and then multiply by the square root of $\hat{v}$. This can be used to construct confidence intervals and t statistics for $\hat{\beta}_1$.

It is useful to see what $\hat{v}$ looks like in some simple cases. When $g = 1$,

$$\hat{v} = \sum_{t=1}^{n} \hat{a}_t^2 + \sum_{t=2}^{n} \hat{a}_t \hat{a}_{t-1}, \tag{12.44}$$

and when $g = 2$,

$$\hat{v} = \sum_{t=1}^{n} \hat{a}_t^2 + (4/3) \left(\sum_{t=2}^{n} \hat{a}_t \hat{a}_{t-1} \right) + (2/3) \left(\sum_{t=3}^{n} \hat{a}_t \hat{a}_{t-2} \right). \tag{12.45}$$

The larger that g is, the more terms are included to correct for serial correlation. The purpose of the factor $[1 - h/(g+1)]$ in (12.42) is to ensure that $\hat{v}$ is in fact nonnegative [Newey and West (1987) verify this]. We clearly need $\hat{v} \geq 0$, since $\hat{v}$ is estimating a variance and the square root of $\hat{v}$ appears in (12.43).

The standard error in (12.43) also turns out to be robust to arbitrary heteroskedasticity. In fact, if we drop the second term in (12.42), then (12.43) becomes the usual heteroskedasticity-robust standard error that we discussed in Chapter 8 (without the degrees of freedom adjustment).

The theory underlying the standard error in (12.43) is technical and somewhat subtle. Remember, we started off by claiming we do not know the form of serial correlation. If this is the case, how can we select the integer g? Theory states that (12.43) works for fairly arbitrary forms of serial correlation, provided g grows with sample size n. The idea is that, with larger sample sizes, we can be more flexible about the amount of correlation in (12.42). There has been much recent work on the relationship between g and n, but we will not go into that here. For annual data, choosing a small g, such as $g = 1$ or $g = 2$, is likely to account for most of the serial correlation. For quarterly or monthly data, g should probably be larger (such as $g = 4$ or 8 for quarterly, $g = 12$ or 24 for monthly), assuming that we have enough data. Newey and West (1987) recommend taking g to be the integer part of $4(n/100)^{2/9}$; others have suggested the integer part of $n^{1/4}$. The Newey-West suggestion is implemented by the econometrics program Eviews®. For, say, $n = 50$ (which is reasonable for annual, postwar data from World War II), $g = 3$. (The integer part of $n^{1/4}$ gives $g = 2$.)

We summarize how to obtain a serial correlation-robust standard error for $\hat{\beta}_1$. Of course, since we can list any independent variable first, the following procedure works for computing a standard error for any slope coefficient.

SERIAL CORRELATION-ROBUST STANDARD ERROR FOR $\hat{\beta}_1$:

(i) Estimate (12.39) by OLS, which yields "se($\hat{\beta}_1$)", $\hat{\sigma}$, and the OLS residuals $\{\hat{u}_t: t = 1, \ldots, n\}$.

(ii) Compute the residuals $\{\hat{r}_t: t = 1, \ldots, n\}$ from the auxiliary regression (12.41). Then form $\hat{a}_t = \hat{r}_t\hat{u}_t$ (for each t).

(iii) For your choice of g, compute $\hat{v}$ as in (12.42).

(iv) Compute se($\hat{\beta}_1$) from (12.43).

Empirically, the serial correlation-robust standard errors are typically larger than the usual OLS standard errors when there is serial correlation. This is because, in most cases, the errors are positively serially correlated. However, it is possible to have substantial serial correlation in $\{u_t\}$ but to also have similarities in the usual and SC-robust standard errors of some coefficients: it is the sample autocorrelations of $\hat{a}_t = \hat{r}_t\hat{u}_t$ that determine the robust standard error for $\hat{\beta}_1$.

The use of SC-robust standard errors has lagged behind the use of standard errors robust only to heteroskedasticity for several reasons. First, large cross sections, where the heteroskedasticity-robust standard errors will have good properties, are more common than large time series. The SC-robust standard errors can be poorly behaved when there is substantial serial correlation and the sample size is small. (Where small can even be as large as, say, 100.) Second, since we must choose the integer g in equation (12.42), computation of the SC-robust standard errors is not automatic. As mentioned earlier, some econometrics packages have automated the selection, but you still have to abide by the choice.

Another important reason that SC-robust standard errors are not yet routinely computed is that, in the presence of severe serial correlation, OLS can be very inefficient, especially in small sample sizes. After performing OLS and correcting the standard errors for serial correlation, the coefficients are often insignificant, or at least less significant than they were with the usual OLS standard errors.

The SC-robust standard errors after OLS estimation are most useful when we have doubts about some of the explanatory variables being strictly exogenous, so that methods such as Cochrane-Orcutt are not even consistent. It is also valid to use the SC-robust standard errors in models with lagged dependent variables assuming, of course, that there is good reason for allowing serial correlation in such models.

EXAMPLE 12.7
(The Puerto Rican Minimum Wage)

We obtain an SC-robust standard error for the minimum wage effect in the Puerto Rican employment equation. In Example 12.2, we found pretty strong evidence of AR(1) serial correlation. As in that example, we use as additional controls log(*usgnp*), log(*prgnp*), and a linear time trend.

The OLS estimate of the elasticity of the employment rate with respect to the minimum wage is $\hat{\beta}_1 = -.2123$, and the usual OLS standard error is "se($\hat{\beta}_1$)" = .0402. The standard error of the regression is $\hat{\sigma} = .0328$. Further, using the previous procedure with $g = 2$ [see (12.45)], we obtain $\hat{v} = .000805$. This gives the SC/heteroskedasticity-robust standard error as $\text{se}(\hat{\beta}_1) = [(.0402/.0328)^2]\sqrt{.000805} \approx .0426$. Interestingly, the robust standard error is only slightly greater than the usual OLS standard error. The robust *t* statistic is about -4.98, and so the estimated elasticity is still very statistically significant.

For comparison, the iterated CO estimate of β_1 is $-.1111$, with a standard error of .0446. Thus, the FGLS estimate is much closer to zero than the OLS estimate, and we might suspect violation of the strict exogeneity assumption. Or, the difference in the OLS and FGLS estimates might be explainable by sampling error. It is very difficult to tell.

Before leaving this section, we note that it is possible to construct serial correlation-robust, *F*-type statistics for testing multiple hypotheses, but these are too advanced to cover here. [See Wooldridge (1991b, 1995) and Davidson and MacKinnon (1993) for treatments.]

12.6 HETEROSKEDASTICITY IN TIME SERIES REGRESSIONS

We discussed testing and correcting for heteroskedasticity for cross-sectional applications in Chapter 8. Heteroskedasticity can also occur in time series regression models, and the presence of heteroskedasticity, while not causing bias or inconsistency in the $\hat{\beta}_j$, does invalidate the usual standard errors, *t* statistics, and *F* statistics. This is just as in the cross-sectional case.

In time series regression applications, heteroskedasticity often receives little, if any, attention: the problem of serially correlated errors is usually more pressing. Nevertheless, it is useful to briefly cover some of the issues that arise in applying tests and corrections for heteroskedasticity in time series regressions.

Since the usual OLS statistics are asymptotically valid under Assumptions TS.1′ through TS.5′, we are interested in what happens when the homoskedasticity assumption, TS.4′, does not hold. Assumption TS.2′ rules out misspecifications such as omitted variables and certain kinds of measurement error, while TS.5′ rules out serial correlation in the errors. It is important to remember that serially correlated errors cause problems which tests and adjustments for heteroskedasticity are not able to address.

Heteroskedasticity-Robust Statistics

In studying heteroskedasticity for cross-sectional regressions, we noted how it has no bearing on the unbiasedness or consistency of the OLS estimators. Exactly the same conclusions hold in the time series case, as we can see by reviewing the assumptions needed for unbiasedness (Theorem 10.1) and consistency (Theorem 11.1).

In Section 8.2, we discussed how the usual OLS standard errors, t statistics, and F statistics can be adjusted to allow for the presence of heteroskedasticity of unknown form. These same adjustments work for time series regressions under Assumptions TS.1′, TS.2′, TS.3′, and TS.5′. Thus, provided the only assumption violated is the homoskedasticity assumption, valid inference is easily obtained in most econometric packages.

Testing for Heteroskedasticity

Sometimes, we wish to test for heteroskedasticity in time series regressions, especially if we are concerned about the performance of heteroskedasticity-robust statistics in relatively small sample sizes. The tests we covered in Chapter 8 can be applied directly, but with a few caveats. First, the errors u_t should *not* be serially correlated; any serial correlation will generally invalidate a test for heteroskedasticity. Thus, it makes sense to test for serial correlation first, using a heteroskedasticity-robust test if heteroskedasticity is suspected. Then, after something has been done to correct for serial correlation, we can test for heteroskedasticity.

Second, consider the equation used to motivate the Breusch-Pagan test for heteroskedasticity:

$$u_t^2 = \delta_0 + \delta_1 x_{t1} + \dots + \delta_k x_{tk} + v_t, \tag{12.46}$$

where the null hypothesis is H_0: $\delta_1 = \delta_2 = \dots = \delta_k = 0$. For the F statistic—with $\hat{u}_t^2$ replacing u_t^2 as the dependent variable—to be valid, we must assume that the errors $\{v_t\}$ are themselves homoskedastic (as in the cross-sectional case) *and* serially uncorrelated. These are implicitly assumed in computing all standard tests for heteroskedasticity, including the version of the White test we covered in Section 8.3. Assuming that the $\{v_t\}$ are serially uncorrelated rules out certain forms of dynamic heteroskedasticity, something we will treat in the next subsection.

If heteroskedasticity is found in the u_t (and the u_t are not serially correlated), then the heteroskedasticity-robust test statistics can be used. An alternative is to use **weighted least squares**, as in Section 8.4. The mechanics of weighted least squares for the time series case are identical to those for the cross-sectional case.

EXAMPLE 12.8

(Heteroskedasticity and the Efficient Markets Hypothesis)

In Example 11.4, we estimated the simple model

$$return_t = \beta_0 + \beta_1 return_{t-1} + u_t. \tag{12.47}$$

The EMH states that $\beta_1 = 0$. When we tested this hypothesis using the data in NYSE.RAW, we obtained $t_{\beta_1} = 1.55$ with $n = 689$. With such a large sample, this is not much evidence against the EMH. While the EMH states that the expected return given past observable information should be constant, it says nothing about the conditional variance. In fact, the Breusch-Pagan test for heteroskedasticity entails regressing the squared OLS residuals $\hat{u}_t^2$ on $return_{t-1}$:

QUESTION 12.5

How would you compute the White test for heteroskedasticity in equation (12.47)?

$$\begin{aligned} \hat{u}_t^2 &= \underset{(0.43)}{4.66} - \underset{(0.201)}{1.104}\, return_{t-1} + residual_t \\ n &= 689,\ R^2 = .042. \end{aligned} \tag{12.48}$$

The t statistic on $return_{t-1}$ is about -5.5, indicating strong evidence of heteroskedasticity. Because the coefficient on $return_{t-1}$ is negative, we have the interesting finding that volatility in stock returns is lower when the previous return was high, and vice versa. Therefore, we have found what is common in many financial studies: the expected value of stock returns does not depend on past returns, but the variance of returns does.

Autoregressive Conditional Heteroskedasticity

In recent years, economists have become interested in dynamic forms of heteroskedasticity. Of course, if $\boldsymbol{x}_t$ contains a lagged dependent variable, then heteroskedasticity as in (12.46) is dynamic. But dynamic forms of heteroskedasticity can appear even in models with no dynamics in the regression equation.

To see this, consider a simple static regression model:

$$y_t = \beta_0 + \beta_1 z_t + u_t,$$

and assume that the Gauss-Markov assumptions hold. This means that the OLS estimators are BLUE. The homoskedasticity assumption says that $\mathrm{Var}(u_t|\boldsymbol{Z})$ is constant, where $\boldsymbol{Z}$ denotes all n outcomes of z_t. Even if the variance of u_t given $\boldsymbol{Z}$ is constant, there are other ways that heteroskedasticity can arise. Engle (1982) suggested looking at the conditional variance of u_t given past errors (where the conditioning on $\boldsymbol{Z}$ is left implicit). Engle suggested what is known as the **autoregressive conditional heteroskedasticity (ARCH)** model. The first order ARCH model is

$$\mathrm{E}(u_t^2|u_{t-1},u_{t-2},\ldots) = \mathrm{E}(u_t^2|u_{t-1}) = \alpha_0 + \alpha_1 u_{t-1}^2, \quad \textbf{(12.49)}$$

where we leave the conditioning on $\boldsymbol{Z}$ implicit. This equation represents the conditional variance of u_t given past u_t, only if $\mathrm{E}(u_t|u_{t-1},u_{t-2},\ldots) = 0$, which means that the errors are serially uncorrelated. Since conditional variances must be positive, this model only makes sense if $\alpha_0 > 0$ and $\alpha_1 \geq 0$; if $\alpha_1 = 0$, there are no dynamics in the variance equation.

It is instructive to write (12.49) as

$$u_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + v_t, \quad \textbf{(12.50)}$$

where the expected value of v_t (given u_{t-1}, u_{t-2}, …) is zero by definition. (The v_t are not independent of past u_t because of the constraint $v_t \geq -\alpha_0 - \alpha_1 u_{t-1}^2$.) Equation (12.50) looks like an autoregressive model in u_t^2 (hence the name ARCH). The stability condition for this equation is $\alpha_1 < 1$, just as in the usual AR(1) model. When $\alpha_1 > 0$, the squared errors contain (positive) serial correlation even though the u_t themselves do not.

What implications does (12.50) have for OLS? Since we began by assuming the Gauss-Markov assumptions hold, OLS is BLUE. Further, even if u_t is not normally distributed, we know that the usual OLS test statistics are asymptotically valid under Assumptions TS.1′ through TS.5′, which are satisfied by static and distributed lag models with ARCH errors.

If OLS still has desirable properties under ARCH, why should we care about ARCH forms of heteroskedasticity in static and distributed lag models? We should be concerned for two reasons. First, it is possible to get consistent (but not unbiased) estimators of the β_j that are *asymptotically* more efficient than the OLS estimators. A weighted least squares procedure, based on estimating (12.50), will do the trick. A maximum likelihood procedure also works under the assumption that the errors u_t have a conditional normal distribution. Second, economists in various fields have become interested in dynamics in the conditional variance. Engle's orginal application was to the variance of United Kingdom inflation, where he found that a larger magnitude of the error in the previous time period (larger u_{t-1}^2) was associated with a larger error variance in the current period. Since variance is often used to measure volatility, and volatility is a key element in asset pricing theories, ARCH models have become important in empirical finance.

ARCH models also apply when there are dynamics in the conditional mean. Suppose we have the dependent variable, y_t, a contemporaneous exogenous variable, z_t, and

$$\mathrm{E}(y_t|z_t,y_{t-1},z_{t-1},y_{t-2},\ldots) = \beta_0 + \beta_1 z_t + \beta_2 y_{t-1} + \beta_3 z_{t-1},$$

so that at most one lag of y and z appears in the dynamic regression. The typical approach is to assume that $\mathrm{Var}(y_t|z_t,y_{t-1},z_{t-1},y_{t-2},\ldots)$ is constant, as we discussed in Chapter 11. But this variance could follow an ARCH model:

$$\begin{aligned}\mathrm{Var}(y_t|z_t,y_{t-1},z_{t-1},y_{t-2},\ldots) &= \mathrm{Var}(u_t|z_t,y_{t-1},z_{t-1},y_{t-2},\ldots)\\ &= \alpha_0 + \alpha_1 u_{t-1}^2,\end{aligned}$$

where $u_t = y_t - \mathrm{E}(y_t|z_t,y_{t-1},z_{t-1},y_{t-2}, \dots)$. As we know from Chapter 11, the presence of ARCH does not affect consistency of OLS, and the usual heteroskedasticity-robust standard errors and test statistics are valid. (Remember, these are valid for any form of heteroskedasticity, and ARCH is just one particular form of heteroskedasticity.)

If you are interested in the ARCH model and its extensions, see Bollerslev, Chou, and Kroner (1992) and Bollerslev, Engle, and Nelson (1994) for recent surveys.

EXAMPLE 12.9
(ARCH in Stock Returns)

In Example 12.8, we saw that there was heteroskedasticity in weekly stock returns. This heteroskedasticity is actually better characterized by the ARCH model in (12.50). If we compute the OLS residuals from (12.47), square these, and regress them on the lagged squared residual, we obtain

$$\hat{u}_t^2 = \underset{(0.44)}{2.95} + \underset{(.036)}{.337}\ \hat{u}_{t-1}^2 + residual_t$$
$$n = 688,\ R^2 = .114. \tag{12.51}$$

The t statistic on $\hat{u}_{t-1}^2$ is over nine, indicating strong ARCH. As we discussed earlier, a larger error at time $t - 1$ implies a larger variance in stock returns today.

It is important to see that, while the *squared* OLS residuals are autocorrelated, the OLS residuals themselves are not (as is consistent with the EMH). Regressing $\hat{u}_t$ on $\hat{u}_{t-1}$ gives $\hat{\rho} = .0014$ with $t_{\hat{\rho}} = .038$.

Heteroskedasticity and Serial Correlation in Regression Models

Nothing rules out the possibility of both heteroskedasticity and serial correlation being present in a regression model. If we are unsure, we can always use OLS and compute fully robust standard errors, as described in Section 12.5.

Much of the time serial correlation is viewed as the most important problem, because it usually has a larger impact on standard errors and the efficiency of estimators than does heteroskedasticity. As we concluded in Section 12.2, obtaining tests for serial correlation that are robust to arbitrary heteroskedasticity is fairly straightforward. If we detect serial correlation using such a test, we can employ the Cochrane-Orcutt transformation [see equation (12.32)] and, in the transformed equation, use heteroskedasticity-robust standard errors and test statistics. Or, we can even test for heteroskedasticity in (12.32) using the Breusch-Pagan or White tests.

Alternatively, we can model heteroskedasticity and serial correlation, and correct for both through a combined weighted least squares AR(1) procedure. Specifically, consider the model

$$\begin{aligned} y_t &= \beta_0 + \beta_1 x_{t1} + \ldots + \beta_k x_{tk} + u_t \\ u_t &= \sqrt{h_t} v_t \\ v_t &= \rho v_{t-1} + e_t, |\rho| < 1, \end{aligned} \quad \textbf{(12.52)}$$

where the explanatory variables $\boldsymbol{X}$ are independent of e_t for all t, and h_t is a function of the x_{tj}. The process $\{e_t\}$ has zero mean, constant variance σ_e^2, and is serially uncorrelated. Therefore, $\{v_t\}$ satisfies a stable AR(1) process. Suppressing the conditioning on the explanatory variables, we have

$$\text{Var}(u_t) = \sigma_v^2 h_t,$$

where $\sigma_v^2 = \sigma_e^2/(1 - \rho^2)$. But $v_t = u_t/\sqrt{h_t}$ is homoskedastic and follows a stable AR(1) model. Therefore, the transformed equation

$$y_t/\sqrt{h_t} = \beta_0(1/\sqrt{h_t}) + \beta_1(x_{t1}/\sqrt{h_t}) + \ldots + \beta_k(x_{tk}/\sqrt{h_t}) + v_t \quad \textbf{(12.53)}$$

has AR(1) errors. Now, if we have a particular kind of heteroskedasticity in mind—that is, we know h_t—we can estimate (12.52) using standard CO or PW methods.

In most cases, we have to estimate h_t first. The following method combines the weighted least squares method from Section 8.4 with the AR(1) serial correlation correction from Section 12.3.

FEASIBLE GLS WITH HETEROSKEDASTICITY AND AR(1) SERIAL CORRELATION:

(i) Estimate (12.52) by OLS and save the residuals, $\hat{u}_t$.
(ii) Regress $\log(\hat{u}_t^2)$ on $x_{t1}, \ldots, x_{tk}$ (or on $\hat{y}_t$, $\hat{y}_t^2$) and obtain the fitted values, say $\hat{g}_t$.
(iii) Obtain the estimates of h_t: $\hat{h}_t = \exp(\hat{g}_t)$.
(iv) Estimate the transformed equation

$$\hat{h}_t^{-1/2} y_t = \hat{h}_t^{-1/2}\beta_0 + \beta_1 \hat{h}_t^{-1/2} x_{t1} + \ldots + \beta_k \hat{h}_t^{-1/2} x_{tk} + error_t \quad \textbf{(12.54)}$$

by standard Cochrane-Orcutt or Prais-Winsten methods.

These feasible GLS estimators are asymptotically efficient. More importantly, all standard errors and test statistics from the CO or PW methods are asymptotically valid.

SUMMARY

We have covered the important problem of serial correlation in the errors of multiple regression models. Positive correlation between adjacent errors is common, especially in static and finite distributed lag models. This causes the usual OLS standard errors and statistics to be misleading (although the $\hat{\beta}_j$ can still be unbiased, or at least consistent). Typically, the OLS standard errors underestimate the true uncertainty in the parameter estimates.

The most popular model of serial correlation is the AR(1) model. Using this as the starting point, it is easy to test for the presence of AR(1) serial correlation using the

OLS residuals. An asymptotically valid t statistic is obtained by regressing the OLS residuals on the lagged residuals, assuming the regressors are strictly exogenous and a homoskedasticity assumption holds. Making the test robust to heteroskedasticity is simple. The Durbin-Watson statistic is available under the classical linear model assumptions, but it can lead to an inconclusive outcome, and it has little to offer over the t test.

For models with a lagged dependent variable, or other nonstrictly exogenous regressors, the standard t test on $\hat{u}_{t-1}$ is still valid, provided all independent variables are included as regressors along with $\hat{u}_{t-1}$. We can use an F or an LM statistic to test for higher order serial correlation.

In models with strictly exogenous regressors, we can use a feasible GLS procedure—Cochrane-Orcutt or Prais-Winsten—to correct for AR(1) serial correlation. This gives estimates that are different from the OLS estimates: the FGLS estimates are obtained from OLS on *quasi-differenced* variables. All of the usual test statistics from the transformed equation are asymptotically valid. Almost all regression packages have built-in features for estimating models with AR(1) errors.

Another way to deal with serial correlation, especially when the strict exogeneity assumption might fail, is to use OLS but to compute serial correlation-robust standard errors (that are also robust to heteroskedasticity). Many regression packages follow a method suggested by Newey and West (1987); it is also possible to use standard regression packages to obtain one standard error at a time.

Finally, we discussed some special features of heteroskedasticity in time series models. As in the cross-sectional case, the most important kind of heteroskedasticity is that which depends on the explanatory variables; this is what determines whether the usual OLS statistics are valid. The Breusch-Pagan and White tests covered in Chapter 8 can be applied directly, with the caveat that the errors should not be serially correlated. In recent years, economists—especially those who study the financial markets—have become interested in dynamic forms of heteroskedasticity. The ARCH model is the leading example.

KEY TERMS

Autoregressive Conditional Heteroskedasticity (ARCH)
Breusch-Godfrey Test
Cochrane-Orcutt (CO) Estimation
Durbin-Watson (*DW*) Statistic
Feasible GLS (FGLS)
Prais-Winsten (PW) Estimation
Quasi-Differenced Data
Serial Correlation-Robust Standard Error
Weighted Least Squares

PROBLEMS

12.1 When the errors in a regression model have AR(1) serial correlation, why do the OLS standard errors tend to underestimate the sampling variation in the $\hat{\beta}_j$? Is it always true that the OLS standard errors are too small?

12.2 Explain what is wrong with the following statement: "The Cochrane-Orcutt and Prais-Winsten methods are both used to obtain valid standard errors for the OLS estimates."

12.3 In Example 10.6, we estimated a variant on Fair's model for predicting presidential election outcomes in the United States.

(i) What argument can be made for the error term in this equation being serially uncorrelated. (*Hint*: How often do presidential elections take place?)

(ii) When the OLS residuals from (10.23) are regressed on the lagged residuals, we obtain $\hat{\rho} = -.068$ and $\text{se}(\hat{\rho}) = .240$. What do you conclude about serial correlation in the u_t?

(iii) Does the small sample size in this application worry you in testing for serial correlation?

12.4 True or False: "If the errors in a regression model contain ARCH, they must be serially correlated."

12.5 (i) In the enterprise zone event study in Problem 10.11, a regression of the OLS residuals on the lagged residuals produces $\hat{\rho} = .841$ and $\text{se}(\hat{\rho}) = .053$. What implications does this have for OLS?

(ii) If you want to use OLS but also want to obtain a valid standard error for the EZ coefficient, what would you do?

12.6 In Example 12.8, we found evidence of heteroskedasticity in u_t in equation (12.47). Thus, we compute the heteroskedasticity-robust standard errors (in $[\cdot]$) along with the usual standard errors:

$$\begin{aligned} \widehat{return}_t = {} & \underset{\substack{(.081) \\ [.085]}}{.180} + \underset{\substack{(.038) \\ [.069]}}{.059}\, return_{t-1} \\ n = 689,\ & R^2 = .0035,\ \bar{R}^2 = .0020. \end{aligned}$$

What does using the heteroskedasticity-robust t statistic do to the significance of $return_{t-1}$?

COMPUTER EXERCISES

12.7 In Example 11.6, we estimated a finite DL model in first differences:

$$\Delta gfr_t = \gamma_0 + \delta_0 \Delta pe_t + \delta_1 \Delta pe_{t-1} + \delta_2 \Delta pe_{t-2} + u_t.$$

Use the data in FERTIL3.RAW to test whether there is AR(1) serial correlation in the errors.

12.8 (i) Using the data in WAGEPRC.RAW, estimate the distributed lag model from Problem 11.5. Use regression (12.14) to test for AR(1) serial correlation.

(ii) Reestimate the model using iterated Cochrane-Orcutt estimation. What is your new estimate of the long-run propensity?

(iii) Using iterated CO, find the standard error for the LRP. (This requires you to estimate a modified equation.) Determine whether the estimated LRP is statistically different from one at the 5% level.

12.9 (i) In part (i) of Problem 11.13, you were asked to estimate the accelerator model for inventory investment. Test this equation for AR(1) serial correlation.

(ii) If you find evidence of serial correlation, reestimate the equation by Cochrane-Orcutt and compare the results.

12.10 (i) Use NYSE.RAW to estimate equation (12.48). Let $\hat{h}_t$ be the fitted values from this equation (the estimates of the conditional variance). How many $\hat{h}_t$ are negative?

(ii) Add $return^2_{t-1}$ to (12.48) and again compute the fitted values, $\hat{h}_t$. Are any $\hat{h}_t$ negative?

(iii) Use the $\hat{h}_t$ from part (ii) to estimate (12.47) by weighted least squares (as in Section 8.4). Compare your estimate of β_1 with that in equation (11.16). Test H_0: $\beta_1 = 0$ and compare the outcome when OLS is used.

(iv) Now, estimate (12.47) by WLS, using the estimated ARCH model in (12.51) to obtain the $\hat{h}_t$. Does this change your findings from part (iii)?

12.11 Consider the version of Fair's model in Example 10.6. Now, rather than predicting the proportion of the two-party vote received by the Democrat, estimate a linear probability model for whether or not the Democrat wins.

(i) Use the binary variable *demwins* in place of *demvote* in (10.23) and report the results in standard form. Which factors affect the probability of winning? Use the data only through 1992.

(ii) How many fitted values are less than zero? How many are greater than one?

(iii) Use the following prediction rule: if $\widehat{demwins} > .5$, you predict the Democrat wins; otherwise, the Republican wins. Using this rule, determine how many of the 20 elections are correctly predicted by the model.

(iv) Plug in the values of the explanatory variables for 1996. What is the predicted probability that Clinton would win the election? Clinton did win; did you get the correct prediction?

(v) Use a heteroskedasticity-robust *t* test for AR(1) serial correlation in the errors. What do you find?

(vi) Obtain the heteroskedasticity-robust standard errors for the estimates in part (i). Are there notable changes in any *t* statistics?

12.12 (i) In Problem 10.13, you estimated a simple relationship between consumption growth and growth in disposable income. Test the equation for AR(1) serial correlation (using CONSUMP.RAW).

(ii) In Problem 11.14, you tested the permanent income hypothesis by regressing the growth in consumption on one lag. After running this regression, test for heteroskedasticity by regressing the squared residuals on gc_{t-1} and gc^2_{t-1}. What do you conclude?