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 $\{x_t: t = 1,2, \ldots\}$ is *stationary* if for every collection of time indices $1 \leq t_1 < t_2 < \ldots < t_m$, the joint distribution of $(x_{t_1}, x_{t_2}, \ldots, x_{t_m})$ is the same as the joint distribution of $(x_{t_1+h}, x_{t_2+h}, \ldots, x_{t_m+h})$ 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 $t = 2,3, \ldots$. In other words, the sequence $\{x_t: t = 1,2, \ldots\}$ 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 $\{x_t: t = 1,2, \ldots\}$ 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 $\{x_t: t = 1,2, \ldots\}$ with finite second moment $[E(x_t^2) < \infty]$ is **covariance stationary** if (i) $E(x_t)$ is constant; (ii) $\mathrm{Var}(x_t)$ is constant; (iii) for any t, $h \geq 1$, $\mathrm{Cov}(x_t,x_{t+h})$ 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.

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

QUESTION 11.1

Suppose that $\{y_t: t = 1,2,\ldots\}$ is generated by $y_t = \delta_0 + \delta_1 t + e_t$, where $\delta_1 \neq 0$, and $\{e_t: t = 1,2,\ldots\}$ 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?

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:

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

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

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

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

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

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 $h \geq 1$:

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

In particular, $\mathrm{Corr}(y_t, y_{t+1}) = \rho_1$, 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 $|\rho_1| < 1$, $\rho_1^h \rightarrow 0$ as $h \rightarrow \infty$. 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, $\mathrm{Corr}(y_t, y_{t+5}) = .591$, $\mathrm{Corr}(y_t, y_{t+10}) = .349$, and $\mathrm{Corr}(y_t, y_{t+20}) = .122$. 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:

$$y_t = \beta_0 + \beta_1 z_{t1} + \beta_2 z_{t2} + u_t. \quad \textbf{(11.7)}$$

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

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

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

$$z_{t1} = \delta_0 + \delta_1 y_{t-1} + v_t. \quad \textbf{(11.9)}$$

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,

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

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

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

This means that, once z_t, z_{t-1}, and z_{t-2} are included, no further lags of z affect $\mathrm{E}(y_t|z_t,z_{t-1},z_{t-2},z_{t-3},\ldots)$; 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:

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

Combined, these two equations imply that

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

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

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

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

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

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

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

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:

$$\text{H}_0\text{: } \beta_1 = \beta_2 = 0. \tag{11.18}$$

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

$$\begin{aligned}\widehat{return}_t &= \underset{(.081)}{.186} + \underset{(.038)}{.060}\ return_{t-1} - \underset{(.038)}{.038}\ return_{t-2} \\ n &= 688,\ R^2 = .0048,\ \bar{R}^2 = .0019\end{aligned}$$

(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

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

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

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

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 $\{e_t\}$,

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

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

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

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.

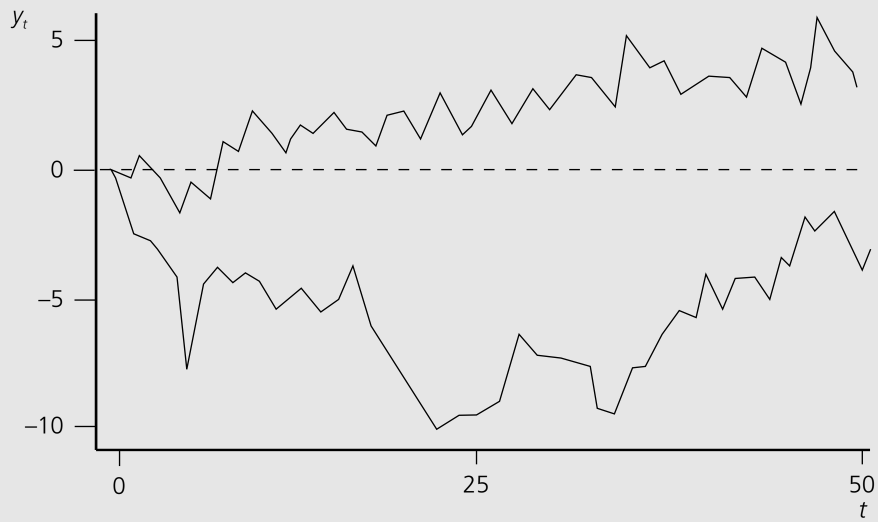
A series that is generally thought to be well-characterized by a random walk is the three-month, T-bill rate. Annual data are plotted in Figure 11.2 for the years 1948 through 1996.

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 $y_0 = 0$, $e_t \sim$ 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**:

$$y_t = \alpha_0 + y_{t-1} + e_t, t = 1,2, \dots, \quad \textbf{(11.23)}$$

where $\{e_t: t = 1,2, \dots\}$ 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:

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

Therefore, if $y_0 = 0$, $\mathrm{E}(y_t) = \alpha_0 t$: 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 $\mathrm{E}(y_{t+h}|y_t) = \alpha_0 h + y_t$, 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.

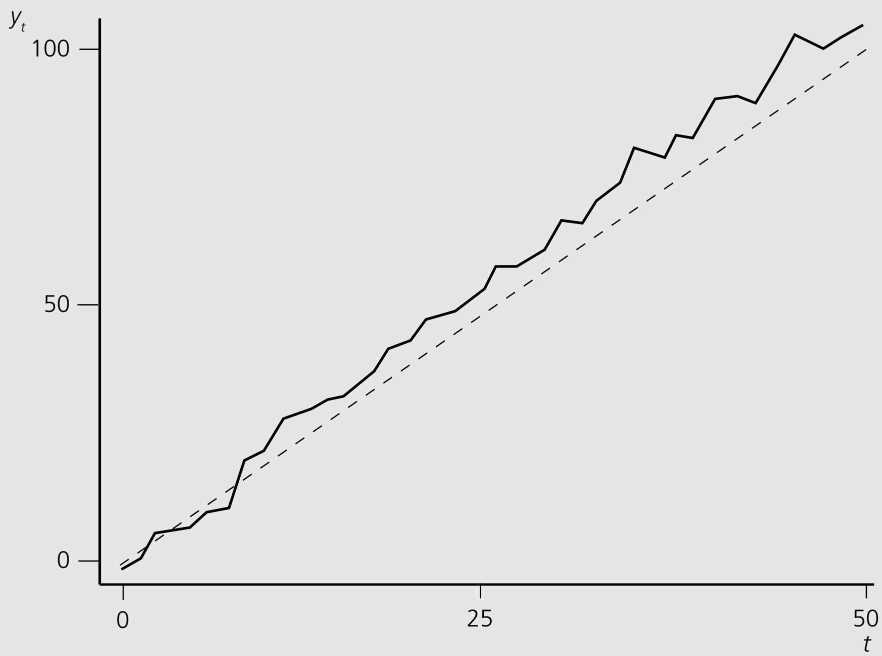
Transformations on Highly Persistent Time Series

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 $t = 1,2, \ldots,$

$$\Delta y_t = y_t - y_{t-1} = e_t, \; t = 2,3, \ldots; \tag{11.24}$$

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

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

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 $\mathrm{E}(\Delta y_t) = \gamma_1 + \mathrm{E}(\Delta v_t) = \gamma_1$. In other words, $\mathrm{E}(\Delta y_t)$ 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, $\rho_1 = \mathrm{Corr}(y_t, y_{t-1})$. 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 $\hat{\rho}_1 > .8$.

E X A M P L E 1 1 . 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):

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

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:

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

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.

E X A M P L E 1 1 . 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

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

(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):

$$\Delta\widehat{\log(hrwage_t)} = \underset{(.0042)}{-.0036} + \underset{(.173)}{.809}\,\Delta\log(outphr)$$
$$n = 40, R^2 = .364, \bar{R}^2 = .348. \qquad \textbf{(11.29)}$$

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

$$y_t = \beta_0 + \beta_1 z_t + u_t, \qquad \textbf{(11.30)}$$

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

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

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

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

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

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

Therefore, even though $\mathrm{E}(y_t|z_t)$ is a linear function of z_t, $\mathrm{Var}(y_t|z_t)$ 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

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

even though $\mathrm{E}(y_t|y_{t-1})$ depends on y_{t-1}, $\mathrm{Var}(y_t|y_{t-1})$ 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

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

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 $\mathrm{Var}(y_t|z_t)$. In equation (11.37), no direct restrictions are placed on, say, $\mathrm{Var}(y_t|y_{t-1})$.

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 $\{y_t: t = 1,2, \ldots\}$ 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

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

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

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

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 Δlog($invpc_t$) 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 H_0: $\beta_1 + \beta_2 = 1$ 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 $\mathrm{E}(return_t|return_{t-1})$ does not depend on $return_{t-1}$. (*Hint*: There are two restrictions to test here.) What do you conclude?

(iii) Drop $return_{t-1}^2$ 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?