Introductory Econometrics
by Jeffrey Wooldridge (2002)
Published by Thomson-South Western

12

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 $E(u_t|\boldsymbol{x}_t) = 0$ 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 $\text{Var}(\hat{\beta}_1)$ in the presence of serial correlation.

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 $\text{Var}(\hat{\beta}_1)$ and show that it is different from the usual formula if $\alpha \neq 0$.

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.

Goodness-of-Fit

Sometimes, one sees the claim that serial correlation in the errors of a time series regression model invalidates our usual goodness-of-fit measures, R-squared, and adjusted R-squared. Fortunately, this is not the case, provided the data are stationary and weakly dependent. To see why these measures are still valid, recall that we defined the population R-squared in a cross-sectional context to be $1—\sigma_u^2/\sigma_y^2$ (see Section 6.3). This definition is still appropriate in the context of time series regressions with stationary, weakly dependent data: the variances of both the errors and the dependent variable do not change over time. By the law of large numbers, R^2 and $\bar{R}^2$ both consistently estimate the population R-squared. The argument is essentially the same as in the cross-sectional case, whether or not there is heteroskedasticity (see Section 8.1). Since there is never an unbiased estimator of the population R-squared, it makes no sense to talk about bias in R^2 caused by serial correlation. All we can really say is that our goodness-of-fit measures are still consistent estimators of the population parameter. This argument does not go through if $\{y_t\}$ is an I(1) process because $\text{Var}(y_t)$ grows with t; goodness-of-fit does not make much sense in this case. As we discussed in Section 10.5, trends in the mean of y_t, or seasonality, can and should be accounted for in computing an R-squared. Other departures from stationarity do not cause difficulty in interepreting R^2 and $\bar{R}^2$ in the usual ways.

Serial Correlation in the Presence of Lagged Dependent Variables

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

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

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

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

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

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

By construction, this model satisfies the key Assumption TS.3′ for consistency of OLS therefore, the OLS estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ are consistent. It is important to see that, with- out further assumptions, the errors $\{u_t\}$ *can* be serially correlated. Condition (12.7) en- sures 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 $-\beta_1 \mathrm{Cov}(u_t,y_{t-2})$, 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

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

Since e_t is uncorrelated with y_{t-1} by assumption, $\mathrm{Cov}(y_{t-1},u_t) = \rho \mathrm{Cov}(y_{t-1},u_{t-1})$, 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 ex- pected 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

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

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—which we will cover 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

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

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

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

and

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

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

$$H_0\colon \rho = 0. \quad (12.12)$$

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

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

(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 $t = 2, \dots, n$, 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 $\hat{\beta}_0$, $\hat{\beta}_1, \dots, \hat{\beta}_k$, 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 $x_{t1}, \dots, x_{tk}$ and obtain the OLS residuals, $\hat{u}_t$, for all $t = 1,2, \dots, n$.

(ii) Run the regression of

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

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 $H_0\colon \rho = 0$ against $H_1\colon \rho \neq 0$ in the usual way. (Actually, since $\rho > 0$ is often expected a priori, the alternative can be $H_1\colon \rho > 0$.) 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

$$H_1: \rho > 0. \tag{12.17}$$

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* statistic 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}; 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} are 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 nonstrictly 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

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

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

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

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

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

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

$$H_0: \rho_1 = 0, \rho_2 = 0, \ldots, \rho_q = 0. \quad \textbf{(12.21)}$$

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

$$u_t = \rho_4 u_{t-4} + e_t. \tag{12.25}$$

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

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

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

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

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

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

where

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

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

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

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

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

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

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

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

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

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 $x_{t1}, \dots, x_{tk}$ and obtain the OLS residuals, $\hat{u}_t$, $t = 1, 2, \dots, n$.

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

(iii) Apply OLS to equation (12.33) to estimate $\beta_0, \beta_1, \dots, \beta_k$. 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.

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	−17.80 (21.05)	−37.31 (23.22)
$\hat{\rho}$	———	.293 (.084)
Observations *R*-Squared	131 .305	130 .193

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 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}.

How can we show that condition (12.35) is needed along with (12.34)? The argument is simple if we assume ρ is known and drop the first time period, as in Cochrane-Orcutt. The argument when we use $\hat{\rho}$ is technically harder and yields no additional insights. Since one observation cannot affect the asymptotic properties of an estimator, dropping it does not affect the argument. Now, with known ρ, the GLS estimator uses $x_t - \rho x_{t-1}$ as the regressor in an equation where $u_t - \rho u_{t-1}$ is the error. From Theorem 11.1, we know the key condition for consistency of OLS is that the error and the regressor are uncorrelated. In this case, we need $\text{E}[(x_t - \rho x_{t-1})(u_t - \rho u_{t-1})] = 0$. If we expand the expectation, we get

$$\begin{aligned}\text{E}[(x_t - \rho x_{t-1})(u_t - \rho u_{t-1})] &= \text{E}(x_t u_t) - \rho\text{E}(x_{t-1}u_t) - \rho\text{E}(x_t u_{t-1}) + \rho^2\text{E}(x_{t-1}u_{t-1})\\ &= -\rho[\text{E}(x_{t-1}u_t) + \text{E}(x_t u_{t-1})]\end{aligned}$$

because $\text{E}(x_t u_t) = \text{E}(x_{t-1}u_{t-1}) = 0$ by assumption (12.34). Now, under stationarity, $\text{E}(x_t u_{t-1}) = \text{E}(x_{t+1}u_t)$ because we are just shifting the time index one period forward. Therefore,

$$\text{E}(x_{t-1}u_t) + \text{E}(x_t u_{t-1}) = \text{E}[(x_{t-1} + x_{t+1})u_t],$$

and the last expectation is the covariance in equation (12.35) because $\text{E}(u_t) = 0$. We have shown that (12.35) is necessary along with (12.34) for GLS to be consistent for β_1. [Of course, if $\rho = 0$, we do not need (12.35) because we are back to doing OLS.]

Our derivation shows 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 it 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 conditions are 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

$$\tilde{y}_t = \beta_0(1 - \rho_1 - \rho_2) + \beta_1\tilde{x}_t + e_t, t = 3,4, \ldots, n. \qquad \textbf{(12.36)}$$

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

$$\hat{u}_t \; on \; \hat{u}_{t-1}, \hat{u}_{t-2}, t = 3, \ldots, n.$$

[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:

$$y_t = \beta_0 + \beta_1 x_t + u_t, t = 1,2, \ldots, \qquad \textbf{(12.37)}$$

where u_t follows the AR(1) process in (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:

$$\Delta y_t = \beta_1 \Delta x_t + \Delta u_t, t = 2, \ldots, n. \qquad \textbf{(12.38)}$$

If u_t follows a random walk, then $e_t \equiv \Delta u_t$ has zero mean and 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), 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

$$y_t = \beta_0 + \beta_1 x_{t1} + \ldots + \beta_k x_{tk} + u_t,\ t=1,2,\ldots,n, \quad (12.39)$$

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,

$$x_{t1} = \delta_0 + \delta_2 x_{t2} + \ldots + \delta_k x_{tk} + r_t, \quad (12.40)$$

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 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

$$x_{t1} \text{ on } x_{t2}, x_{t3}, \ldots, x_{tk} \quad (12.41)$$

(including a constant, as usual). For a chosen integer $g > 0$, define

$$\hat{\nu} = \sum_{t=1}^{n} \hat{a}_t^2 + 2\sum_{h=1}^{g} [1 - h/(g+1)]\left(\sum_{t=h+1}^{n} \hat{a}_t \hat{a}_{t-h}\right), \quad (12.42)$$

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 $\hat{\nu}$, the **serial correlation-robust standard error** of $\hat{\beta}_1$ is simply

$$\text{se}(\hat{\beta}_1) = [\text{“se}(\hat{\beta}_1)\text{”}/\hat{\sigma}]^2\sqrt{\hat{\nu}}. \quad (12.43)$$

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 $\hat{\nu}$. This can be used to construct confidence intervals and t statistics for $\hat{\beta}_1$.

It is useful to see what $\hat{\nu}$ looks like in some simple cases. When $g = 1$,

$$\hat{\nu} = \sum_{t=1}^{n} \hat{a}_t^2 + \sum_{t=2}^{n} \hat{a}_t \hat{a}_{t-1}, \tag{12.44}$$

and when $g = 2$,

$$\hat{\nu} = \sum_{t=1}^{n} \hat{a}_t^2 + (4/3)\left(\sum_{t=2}^{n} \hat{a}_t \hat{a}_{t-1}\right) + (2/3)\left(\sum_{t=3}^{n} \hat{a}_t \hat{a}_{t-2}\right). \tag{12.45}$$

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 $\hat{\nu}$ is in fact nonnegative [Newey and West (1987) verify this]. We clearly need $\hat{\nu} \geq 0$, since $\hat{\nu}$ is estimating a variance and the square root of $\hat{\nu}$ 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 and $g = 12$ 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 $\hat{\nu}$ 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 serial correlation-robust (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:

$$u_t^2 = \delta_0 + \delta_1 x_{t1} + \ldots + \delta_k x_{tk} + v_t, \quad (12.46)$$

where the null hypothesis is H_0: $\delta_1 = \delta_2 = \ldots = \delta_k = 0$. 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

$$return_t = \beta_0 + \beta_1 return_{t-1} + u_t. \quad (12.47)$$

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)?

$$\begin{aligned} \hat{u}_t^2 &= \underset{(0.43)}{4.66} - \underset{(0.201)}{1.104}\, return_{t-1} + residual_t \\ n &= 689,\ R^2 = .042. \end{aligned} \quad (12.48)$$

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 $\mathbf{x}_t$ 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 $\text{Var}(u_t|\mathbf{Z})$ is constant, where $\mathbf{Z}$ denotes all n outcomes of z_t. Even if the variance of u_t given $\mathbf{Z}$ 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 $\mathbf{Z}$ is left implicit). Engle suggested what is known as the **autoregressive conditional heteroskedasticity (ARCH)** model. The first order ARCH model is

$$\text{E}(u_t^2|u_{t-1},u_{t-2},\ldots) = \text{E}(u_t^2|u_{t-1}) = \alpha_0 + \alpha_1 u_{t-1}^2, \quad \textbf{(12.49)}$$

where we leave the conditioning on $\mathbf{Z}$ implicit. This equation represents the conditional variance of u_t given past u_t, only if $\text{E}(u_t|u_{t-1},u_{t-2},\ldots) = 0$, 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 original 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 $u_t = y_t - \mathrm{E}(y_t|z_t,y_{t-1},z_{t-1},y_{t-2},\ldots)$. 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

$$\begin{aligned}\hat{u}_t^2 &= \underset{(.44)}{2.95} + \underset{(.036)}{.337}\,\hat{u}_{t-1}^2 + \mathit{residual},\\ n &= 688,\ R^2 = .114.\end{aligned} \tag{12.51}$$

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

$$\begin{aligned} y_t &= \beta_0 + \beta_1 x_{t1} + \dots + \beta_k x_{tk} + u_t \\ u_t &= \sqrt{h_t}v_t \\ v_t &= \rho v_{t-1} + e_t, \ |\rho| < 1, \end{aligned} \tag{12.52}$$

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 and 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

$$y_t/\sqrt{h_t} = \beta_0(1/\sqrt{h_t}) + \beta_1(x_{t1}/\sqrt{h_t}) + \dots + \beta_k(x_{tk}/\sqrt{h_t}) + v_t \tag{12.53}$$

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 $x_{t1}, \dots, x_{tk}$ (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 [·]) 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 $return_{t-1}^2$ 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.

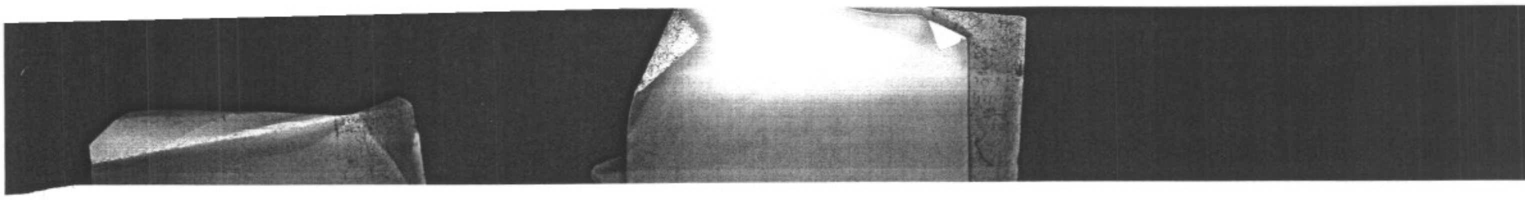
(iv) Now, estimate (12.47) by WLS, using the estimated ARCH model in (12.51) to obtain the $\hat{h}_t$. Does this change your findings from part (iii)?

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, deter-



mine 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_{t-1}^2. What do you conclude?

12.13 (i) For Example 12.4, using the data in BARIUM.RAW, obtain the iterative Prais-Winsten estimates.

(ii) Are the Prais-Winsten and Cochrane-Orcutt estimates similar? Did you expect them to be?

12.14 Use the data in TRAFFIC2.RAW for this exercise.

(i) Run an OLS regression of *prcfat* on a linear time trend, monthly dummy variables, and the variables *wkends*, *unem*, *spdlaw*, and *beltlaw*. Test the errors for AR(1) serial correlation using the regression in equation (12.14). Does it make sense to use the test that assumes strict exogeneity of the regressors?

(ii) Obtain serial correlation- and heteroskedasticity-robust standard errors for the coefficients on *spdlaw* and *bltlaw*, using four lags in the Newey-West estimator. How does this affect the statistical significance of the two policy variables?

(iii) Now, estimate the model using iterative Prais-Winsten and compare the estimates with the OLS estimates. Are there important changes in the policy variable coefficients or their statistical significance?

12.15 The file FISH.RAW contains 97 daily price and quantity observations on fish prices at the Fulton Fish Market in Manhattan. Use the variable log(*avgprc*) as the dependent variable.

(i) Regress log(*avgprc*) on four daily dummy variables, with Friday as the base. Include a linear time trend. Is there evidence that price varies systematically within a week?

(ii) Now, add the variables *wave*2 and *wave*3, which are measures of wave heights over the past several days. Are these variables individually significant? Describe a mechanism by which stormier seas would increase the price of fish.

(iii) What happened to the time trend when *wave*2 and *wave*3 were added to the regression? What must be going on?

(iv) Explain why all explanatory variables in the regression are safely assumed to be strictly exogenous.

(v) Test the errors for AR(1) serial correlation.

(vi) Obtain the Newey-West standard errors using four lags. What happens to the t statistics on *wave*2 and *wave*3? Did you expect a bigger or smaller change compared with the usual OLS t statistics?

(vii) Now, obtain the Prais-Winsten estimates for the model estimated in part (ii). Are *wave*2 and *wave*3 jointly statistically significant?