$[SSR(1) - SSR(2)]/[SSR(2)/(T-2)]$ is the homoskedasticity-only F-statistic (Equation 7.13) testing the null hypothesis that $\beta_2 = 0$ in the AR(2). If u_t is homoskedastic, then F has a χ_1^2 asymptotic distribution; if not, it has some other asymptotic distribution. Thus pr[BIC(2) − BIC(1) < 0] = $\Pr\{T[\text{BIC}(2) - \text{BIC}(1)] < 0\} = \Pr\{-T\ln[1 + F/(T-2)] + (\ln T) < 0\} = \Pr\{T\ln[1 + F/(T-2)] > \ln T\}$. As T increases, $T\ln[1 + F/(T-2)] - F \xrightarrow{p} 0$ [a consequence of the logarithmic approximation $\ln(1 + a) \cong a$, which becomes exact as $a \longrightarrow 0$]. Thus $\Pr[\text{BIC}(2) - \text{BIC}(1) < 0] \longrightarrow \Pr(F > \ln T) \longrightarrow 0$, so $\Pr(\hat{p} = 2) \longrightarrow 0$.

AIC

In the special case of an AR(1) when zero, one, or two lags are considered, (i) applies to the AIC where the term $\ln T$ is replaced by 2, so $\Pr(\hat{p} = 0) \longrightarrow 0$. All the steps in the proof of (ii) for the BIC also apply to the AIC, with the modification that $\ln T$ is replaced by 2; thus $\Pr[\text{AIC}(2) - \text{AIC}(1) < 0] \longrightarrow \Pr(F > 2) > 0$. If u_t is homoskedastic, then $\Pr(F > 2) \longrightarrow \Pr(\chi_1^2 > 2) = 0.16$, so $\Pr(\hat{p} = 2) \longrightarrow 0.16$. In general, when $\hat{p}$ is chosen using the AIC, $\Pr(\hat{p} < p) \longrightarrow 0$ but $\Pr(\hat{p} > p)$ tends to a positive number, so $\Pr(\hat{p} = p)$ does not tend to 1.

CHAPTER 15 Estimation of Dynamic Causal Effects

In the 1983 movie *Trading Places*, the characters played by Dan Aykroyd and Eddie Murphy used inside information on how well Florida oranges had fared over the winter to make millions in the orange juice concentrate futures market, a market for contracts to buy or sell large quantities of orange juice concentrate at a specified price on a future date. In real life, traders in orange juice futures in fact do pay close attention to the weather in Florida: Freezes in Florida kill Florida oranges, the source of almost all frozen orange juice concentrate made in the United States, so its supply falls and the price rises. But precisely how much does the price rise when the weather in Florida turns sour? Does the price rise all at once, or are there delays; if so, for how long? These are questions that real-life traders in orange juice futures need to answer if they want to succeed.

This chapter takes up the problem of estimating the effect on Y now and in the future of a change in X, that is, the **dynamic causal effect** on Y of a change in X. What, for example, is the effect on the path of orange juice prices over time of a freezing spell in Florida? The starting point for modeling and estimating dynamic causal effects is the so-called distributed lag regression model, in which Y_t is expressed as a function of current and past values of X_t. Section 15.1 introduces the distributed lag model in the context of estimating the effect of cold weather in Florida on the price of orange juice concentrate over time. Section 15.2 takes a closer look at what, precisely, is meant by a dynamic causal effect.

One way to estimate dynamic causal effects is to estimate the coefficients of the distributed lag regression model using OLS. As discussed in Section 15.3, this estimator is consistent if the regression error has a conditional mean of zero given current and past values of X, a condition that (as in Chapter 12) is referred to as exogeneity. Because the omitted determinants of Y_t are correlated over time—that is, because they are serially correlated—the error term in the distributed lag model can be serially correlated. This possibility in turn requires "heteroskedasticity- and autocorrelation-consistent" (HAC) standard errors, the topic of Section 15.4.

A second way to estimate dynamic causal effects, discussed in Section 15.5, is to model the serial correlation in the error term as an autoregression and then to use this autoregressive model to derive an autoregressive distributed lag (ADL) model. Alternatively, the coefficients of the original distributed lag model can be estimated by generalized least squares (GLS). Both the ADL and GLS methods, however, require a stronger version of exogeneity than we have used so far: *strict* exogeneity, under which the regression errors have a conditional mean of zero given past, present, *and future* values of X.

Section 15.6 provides a more complete analysis of the relationship between orange juice prices and the weather. In this application, the weather is beyond human control and thus is exogenous (although, as discussed in Section 15.6, economic theory suggests that it is not necessarily strictly exogenous). Because exogeneity is necessary for estimating dynamic causal effects, Section 15.7 examines this assumption in several applications taken from macroeconomics and finance.

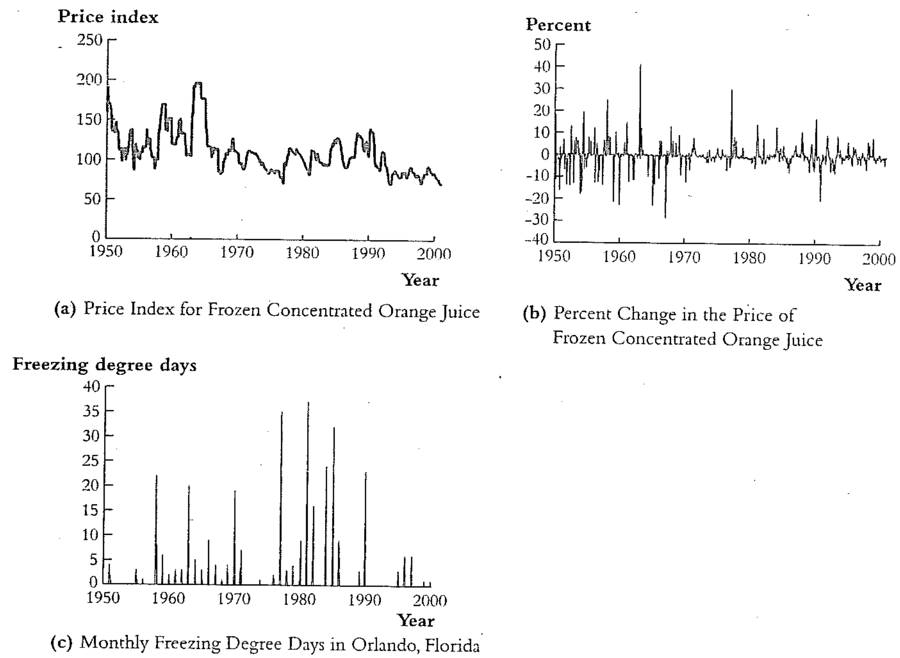
This chapter builds on the material in Sections 14.1 through 14.4 but, with the exception of a subsection (that can be skipped) of the empirical analysis in Section 15.6, does not require the material in Sections 14.5 through 14.7.

15.1 An Initial Taste of the Orange Juice Data

Orlando, the historical center of Florida's orange-growing region, is normally sunny and warm. But now and then there is a cold snap, and if temperatures drop below freezing for too long, the trees drop many of their oranges. If the cold snap is severe, the trees freeze. Following a freeze, the supply of orange juice concentrate falls and its price rises. The timing of the price increases is rather complicated, however. Orange juice concentrate is a "durable," or storable, commodity; that is, it can be stored in its frozen state, albeit at some cost (to run the freezer). Thus the price of orange juice concentrate depends not only on current supply but also on expectations of future supply. A freeze today means that future supplies of concentrate will be low, but because concentrate currently in storage can be used to meet either current or future demand, the price of existing concentrate rises today. But precisely how much does the price of concentrate rise when there is a freeze? The answer to this question is of interest not just to orange juice traders but more generally to economists interested in studying the operations of modern commodity markets. To learn how the price of orange juice changes in response to weather conditions, we must analyze data on orange juice prices and the weather.

Monthly data on the price of frozen orange juice concentrate, its monthly percentage change, and temperatures in the orange-growing region of Florida from January 1950 to December 2000 are plotted in Figure 15.1. The price, plotted in Figure 15.1a, is a measure of the average real price of frozen orange juice concentrate paid by wholesalers. This price was deflated by the overall producer price

FIGURE 15.1 Orange Juice Prices and Florida Weather, 1950–2000



There have been large month-to-month changes in the price of frozen concentrated orange juice. Many of the large movements coincide with freezing weather in Orlando, home of many orange groves.

index for finished goods to eliminate the effects of overall price inflation. The percentage price change plotted in Figure 15.1b is the percent change in the price over the month. The temperature data plotted in Figure 15.1c are the number of "freezing degree days" at the Orlando, Florida, airport, calculated as the sum of the number of degrees Fahrenheit that the minimum temperature falls below freezing in a given day over all days in the month; for example, in November 1950 the airport temperature dropped below freezing twice, on the 25th (31°) and on the 29th (29°), for a total of 4 freezing degree days $[(32 - 31) + (32 - 29) = 4]$. (The data are described in more detail in Appendix 15.1.) As you can see by comparing the panels in Figure 15.1, the price of orange juice concentrate has large swings, some of which appear to be associated with cold weather in Florida.

We begin our quantitative analysis of the relationship between orange juice price and the weather by using a regression to estimate the amount by which orange juice prices rise when the weather turns cold. The dependent variable is the percentage change in the price over that month [$\%ChgP_t$, where $\%ChgP_t = 100 \times \Delta\ln(P_t^{OJ})$ and P_t^{OJ} is the real price of orange juice]. The regressor is the number of freezing degree days during that month (FDD_t). This regression is estimated using monthly data from January 1950 to December 2000 (as are all regressions in this chapter), for a total of $T = 612$ observations:

$$\widehat{\%ChgP_t} = \underset{(0.22)}{-0.40} + \underset{(0.13)}{0.47}FDD_t. \tag{15.1}$$

The standard errors reported in this section are not the usual OLS standard errors, but rather are heteroskedasticity- and autocorrelation-consistent (HAC) standard errors that are appropriate when the error term and regressors are autocorrelated. HAC standard errors are discussed in Section 15.4, and for now they are used without further explanation.

According to this regression, an additional freezing degree day during a month increases the price of orange juice concentrate over that month by 0.47%. In a month with 4 freezing degree days, such as November 1950, the price of orange juice concentrate is estimated to have increased by 1.88% ($4 \times 0.47\% = 1.88\%$), relative to a month with no days below freezing.

Because the regression in Equation (15.1) includes only a contemporaneous measure of the weather, it does not capture any lingering effects of the cold snap on the orange juice price over the coming months. To capture these we need to consider the effect on prices of both contemporaneous and lagged values of FDD, which in turn can be done by augmenting the regression in Equation (15.1) with, for example, lagged values of FDD over the previous 6 months:

$$\begin{aligned}\widehat{\%ChgP_t} = &\underset{(0.23)}{-0.65} + \underset{(0.14)}{0.47}FDD_t + \underset{(0.08)}{0.14}FDD_{t-1} + \underset{(0.06)}{0.06}FDD_{t-2} \\ &+ \underset{(0.05)}{0.07}FDD_{t-3} + \underset{(0.03)}{0.03}FDD_{t-4} + \underset{(0.03)}{0.05}FDD_{t-5} + \underset{(0.04)}{0.05}FDD_{t-6}.\end{aligned} \tag{15.2}$$

Equation (15.2) is a distributed lag regression. The coefficient on FDD_t in Equation (15.2) estimates the percentage increase in prices over the course of the month in which the freeze occurs; an additional freezing degree day is estimated to increase prices that month by 0.47%. The coefficient on the first lag of FDD_t, FDD_{t-1}, estimates the percentage increase in prices arising from a freezing degree day in the preceding month, the coefficient on the second lag estimates the effect of a freezing degree day 2 months ago, and so forth. Equivalently, the coefficient on the first lag of FDD estimates the effect of a unit increase in FDD 1 month after the freeze occurs. Thus the estimated coefficients in Equation (15.2) are estimates of the effect of a unit increase in FDD_t on current and future values of $\%ChgP$; that is, they are estimates of the dynamic effect of FDD_t on $\%ChgP_t$. For example, the 4 freezing degree days in November 1950 are estimated to have increased orange juice prices by 1.88% during November 1950, by an additional 0.56% ($= 4 \times 0.14$) in December 1950, by an additional 0.24% ($= 4 \times 0.06$) in January 1951, and so forth.

15.2 Dynamic Causal Effects

Before learning more about the tools for estimating dynamic causal effects, we should spend a moment thinking about what, precisely, is meant by a dynamic causal effect. Having a clear idea about what a dynamic causal effect is leads to a clearer understanding of the conditions under which it can be estimated.

Causal Effects and Time Series Data

Section 1.2 defined a causal effect as the outcome of an ideal randomized controlled experiment: When a horticulturalist randomly applies fertilizer to some tomato plots but not others and then measures the yield, the expected difference in yield between the fertilized and unfertilized plots is the causal effect on tomato yield of the fertilizer. This concept of an experiment, however, is one in which there are multiple subjects (multiple tomato plots or multiple people), so the data are either cross-sectional (the tomato yield at the end of the harvest) or panel data (individual incomes before and after an experimental job training program). By having multiple subjects, it is possible to have both treatment and control groups and thereby to estimate the causal effect of the treatment.

In time series applications, this definition of causal effects in terms of an ideal randomized controlled experiment needs to be modified. To be concrete, consider an important problem of macroeconomics: estimating the effect of an unanticipated change in the short-term interest rate on the current and future economic activity in a given country, as measured by GDP. Taken literally, the randomized controlled experiment of Section 1.2 would entail randomly assigning different economies to treatment and control groups. The central banks in the treatment group would apply the treatment of a random interest rate change, while those in the control group would apply no such random changes; for both groups, economic activity (for example, GDP) would be measured over the next few years. But what if we are interested in estimating this effect for a specific country, say the United States? Then this experiment would entail having different "clones" of the United States as subjects and assigning some clone economies to the treatment group and some to the control group. Obviously, this "parallel universes" experiment is infeasible.

Instead, in time series data it is useful to think of a randomized controlled experiment consisting of the same subject (e.g., the U.S. economy) being given different treatments (randomly chosen changes in interest rates) at different points in time (the 1970s, the 1980s, and so forth). In this framework, the single subject at different times plays the role of both treatment and control group: Sometimes the Fed changes the interest rate, while at other times it does not. Because data are collected over time, it is possible to estimate the dynamic causal effect, that is, the time path of the effect on the outcome of interest of the treatment. For example, a surprise increase in the short-term interest rate of two percentage points, sustained for one quarter, might initially have a negligible effect on output; after two quarters GDP growth might slow, with the greatest slowdown after $1\frac{1}{2}$ years; then over the next 2 years, GDP growth might return to normal. This time path of causal effects is the dynamic causal effect on GDP growth of a surprise change in the interest rate.

As a second example, consider the causal effect on orange juice price changes of a freezing degree day. It is possible to imagine a variety of hypothetical experiments, each yielding a different causal effect. One experiment would be to change the weather in the Florida orange groves, holding constant weather elsewhere—for example, holding constant weather in the Texas grapefruit groves and in other citrus fruit regions. This experiment would measure a partial effect, holding other weather constant. A second experiment might change the weather in all the regions, where the "treatment" is application of overall weather patterns. If weather is correlated across regions for competing crops, then these two dynamic causal effects differ. In this chapter, we consider the causal effect in the latter experiment, that is, the causal effect of applying general weather patterns. This corresponds to measuring the dynamic effect on prices of a change in Florida weather, *not* holding constant weather in other agricultural regions.

Dynamic effects and the distributed lag model. Because dynamic effects necessarily occur over time, the econometric model used to estimate dynamic causal effects needs to incorporate lags. To do so, Y_t can be expressed as a distributed lag of current and r past values of X_t:

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 X_{t-1} + \beta_3 X_{t-2} + \cdots + \beta_{r+1} X_{t-r} + u_t, \tag{15.3}$$

where u_t is an error term that includes measurement error in Y_t and the effect of omitted determinants of Y_t. The model in Equation (15.3) is called the **distributed lag model** relating X_t, and r of its lags, to Y_t.

As an illustration of Equation (15.3), consider a modified version of the tomato/fertilizer experiment: Because fertilizer applied today might remain in the ground in future years, the horticulturalist wants to determine the effect on tomato yield *over time* of applying fertilizer. Accordingly, she designs a 3-year experiment and randomly divides her plots into four groups: The first is fertilized in only the first year; the second is fertilized in only the second year; the third is fertilized in only the third year; and the fourth, the control group, is never fertilized. Tomatoes are grown annually in each plot, and the third-year harvest is weighed. The three treatment groups are denoted by the binary variables X_{t-2}, X_{t-1}, and X_t, where t represents the third year (the year in which the harvest is weighed), $X_{t-2} = 1$ if the plot is in the first group (fertilized two years earlier), $X_{t-1} = 1$ if the plot was fertilized 1 year earlier, and $X_t = 1$ if the plot was fertilized in the final year. In the context of Equation (15.3) (which applies to a single plot), the effect of being fertilized in the final year is β_1, the effect of being fertilized 1 year earlier is β_2, and the effect of being fertilized 2 years earlier is β_3. If the effect of fertilizer is greatest in the year it is applied, then β_1 would be larger than β_2 and β_3.

More generally, the coefficient on the contemporaneous value of X_t, β_1, is the contemporaneous or immediate effect of a unit change in X_t on Y_t. The coefficient on X_{t-1}, β_2, is the effect on Y_t of a unit change in X_{t-1} or, equivalently, the effect on Y_{t+1} of a unit change in X_t; that is, β_2 is the effect of a unit change in X on Y one period later. In general, the coefficient on X_{t-h} is the effect of a unit change in X on Y after h periods. The dynamic causal effect is the effect of a change in X_t on Y_t, Y_{t+1}, Y_{t+2}, and so forth; that is, it is the sequence of causal effects on current and future values of Y. Thus, in the context of the distributed lag model in Equation (15.3), the dynamic causal effect is the sequence of coefficients β_1, $\beta_2, \ldots, \beta_{r+1}$.

Implications for empirical time series analysis. This formulation of dynamic causal effects in time series data as the expected outcome of an experiment in which different treatment levels are repeatedly applied to the same subject has two implications for empirical attempts to measure the dynamic causal effect with observational time series data. The first implication is that the dynamic causal effect should not change over the sample on which we have data. This in turn is implied by the data being jointly stationary (Key Concept 14.5). As discussed in Section 14.7, the hypothesis that a population regression function is stable over time can be tested using the QLR test for a break, and it is possible to estimate the dynamic causal effect in different subsamples. The second implication is that X must be uncorrelated with the error term, and it is to this implication that we now turn.

Two Types of Exogeneity

Section 12.1 defined an "exogenous" variable as a variable that is uncorrelated with the regression error term and an "endogenous" variable as a variable that is correlated with the error term. This terminology traces to models with multiple equations, in which an "endogenous" variable is determined within the model while an "exogenous" variable is determined outside the model. Loosely speaking, if we are to estimate dynamic causal effects using the distributed lag model in Equation (15.3), the regressors (the X's) must be uncorrelated with the error term. Thus X must be exogenous. Because we are working with time series data, however, we need to refine the definitions of exogeneity. In fact, there are two different concepts of exogeneity that we use here.

The first concept of exogeneity is that the error term has a conditional mean of zero given current and all past values of X_t, that is, that $E(u_t|X_t, X_{t-1}, X_{t-2}, \ldots) = 0$. This modifies the standard conditional mean assumption for multiple regression with cross-sectional data (Assumption #1 in Key Concept 6.4), which requires only that u_t has a conditional mean of zero given the included regressors, that is, that $E(u_t|X_t, X_{t-1}, \ldots, X_{t-r}) = 0$. Including all lagged values of X_t in the conditional expectation implies that all the more distant causal effects—all the causal effects beyond lag r—are zero. Thus, under this assumption, the r distributed lag coefficients in Equation (15.3) constitute all the nonzero dynamic causal effects. We can refer to this assumption—that $E(u_t|X_t, X_{t-1}, \ldots) = 0$—as *past and present exogeneity*, but because of the similarity of this definition and the definition of exogeneity in Chapter 12, we just use the term **exogeneity**.

The second concept of exogeneity is that the error term has mean zero, given all past, present, and *future* values of X_t, that is, that $E(u_t|\ldots, X_{t+2}, X_{t+1}, X_t, X_{t-1}, X_{t-2}, \ldots) = 0$. This is called **strict exogeneity**; for clarity, we also call it *past, present, and future exogeneity*. The reason for introducing the concept of strict exogeneity is that, when X is strictly exogenous, there are more efficient estimators of dynamic causal effects than the OLS estimators of the coefficients of the distributed lag regression in Equation (15.3).

The difference between exogeneity (past and present) and strict exogeneity (past, present, and future) is that strict exogeneity includes future values of X in the conditional expectation. Thus strict exogeneity implies exogeneity, but not the reverse. One way to understand the difference between the two concepts is to consider the implications of these definitions for correlations between X and u. If X is (past and present) exogenous, then u_t is uncorrelated with current and past

The Distributed Lag Model and Exogeneity

KEY CONCEPT 15.1

In the distributed lag model

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 X_{t-1} + \beta_3 X_{t-2} + \cdots + \beta_{r+1} X_{t-r} + u_t, \quad (15.4)$$

there are two different types of exogeneity, that is, two different exogeneity conditions:

Past and present exogeneity (exogeneity):

$$E(u_t|X_t, X_{t-1}, X_{t-2}, \ldots) = 0; \quad (15.5)$$

Past, present, and future exogeneity (strict exogeneity):

$$E(u_t|\ldots, X_{t+2}, X_{t+1}, X_t, X_{t-1}, X_{t-2}, \ldots) = 0. \quad (15.6)$$

If X is strictly exogenous, it is exogenous, but exogeneity does not imply strict exogeneity.

values of X_t. If X is strictly exogenous, then in addition u_t is uncorrelated with *future* values of X_t. For example, if a change in Y_t causes *future* values of X_t to change, then X_t is not strictly exogenous even though it might be (past and present) exogenous.

As an illustration, consider the hypothetical multiyear tomato/fertilizer experiment described following Equation (15.3). Because the fertilizer is randomly applied in the hypothetical experiment, it is exogenous. Because tomato yield today does not depend on the amount of fertilizer applied in the future, the fertilizer time series is also strictly exogenous.

As a second illustration, consider the orange juice price example, in which Y_t is the monthly percentage change in orange juice prices and X_t is the number of freezing degree days in that month. From the perspective of orange juice markets, we can think of the weather—the number of freezing degree days—as if it were randomly assigned, in the sense that the weather is outside human control. If the effect of FDD is linear and if it has no effect on prices after r months, then it follows that the weather is exogenous. But is the weather *strictly* exogenous? If the conditional mean of u_t given future FDD is nonzero, then FDD is not strictly exogenous. Answering this question requires thinking carefully about what, precisely, is contained in u_t. In particular, if OJ market participants use forecasts of FDD when they decide how much they will buy or sell at a given price, then OJ prices, and thus the error term u_t, could incorporate information about future *FDD* that would make u_t a useful predictor of *FDD*. This means that u_t will be correlated with future values of FDD_t. According to this logic, because u_t includes forecasts of future Florida weather, *FDD* would be (past and present) exogenous but not *strictly* exogenous. The difference between this and the tomato/fertilizer example is that, while tomato plants are unaffected by future fertilization, OJ market participants *are* influenced by forecasts of future Florida weather. We return to the question of whether *FDD* is strictly exogenous when we analyze the orange juice price data in more detail in Section 15.6.

The two definitions of exogeneity are summarized in Key Concept 15.1.

15.3 Estimation of Dynamic Causal Effects with Exogenous Regressors

If X is exogenous, then its dynamic causal effect on Y can be estimated by OLS estimation of the distributed lag regression in Equation (15.4). This section summarizes the conditions under which these OLS estimators lead to valid statistical inferences and introduces dynamic multipliers and cumulative dynamic multipliers.

The Distributed Lag Model Assumptions

The four assumptions of the distributed lag regression model are similar to the four assumptions for the cross-sectional multiple regression model (Key Concept 6.4), modified for time series data.

The first assumption is that X is exogenous, which extends the zero conditional mean assumption for cross-sectional data to include all lagged values of X. As discussed in Section 15.2, this assumption implies that the r distributed lag coefficients in Equation (15.3) constitute all the nonzero dynamic causal effects. In this sense, the population regression function summarizes the entire dynamic effect on Y of a change in X.

The second assumption has two parts: Part (a) requires that the variables have a stationary distribution, and part (b) requires that they become independently distributed when the amount of time separating them becomes large. This assumption is the same as the corresponding assumption for the ADL model (the second assumption in Key Concept 14.6), and the discussion of this assumption in Section 14.4 applies here as well.

The third assumption is that large outliers are unlikely, made mathematically precise by assuming that the variables have more than eight nonzero, finite moments.

The Distributed Lag Model Assumptions

KEY CONCEPT 15.2

The distributed lag model is given in Key Concept 15.1 [Equation (15.4)], where

1. X is exogenous, that is, $E(u_t|X_t, X_{t-1}, X_{t-2}, \ldots) = 0$;
2. (a) The random variables Y_t and X_t have a stationary distribution, and
   (b) (Y_t, X_t) and (Y_{t-j}, X_{t-j}) become independent as j gets large;
3. Large outliers are unlikely: Y_t and X_t have more than eight nonzero, finite moments; and
4. There is no perfect multicollinearity.

This is stronger than the assumption of four finite moments that is used elsewhere in this book. As discussed in Section 15.4, this stronger assumption is used in the mathematics behind the HAC variance estimator.

The fourth assumption, which is the same as in the cross-sectional multiple regression model, is that there is no perfect multicollinearity.

The distributed lag regression model and assumptions are summarized in Key Concept 15.2.

Extension to additional X's. The distributed lag model extends directly to multiple X's: The additional X's and their lags are simply included as regressors in the distributed lag regression, and the assumptions in Key Concept 15.2 are modified to include these additional regressors. Although the extension to multiple X's is conceptually straightforward, it complicates the notation, obscuring the main ideas of estimation and inference in the distributed lag model. For this reason, the case of multiple X's is not treated explicitly in this chapter but is left as a straightforward extension of the distributed lag model with a single X.

Autocorrelated u_t, Standard Errors, and Inference

In the distributed lag regression model, the error term u_t can be autocorrelated; that is, u_t can be correlated with its lagged values. This autocorrelation arises because, in time series data, the omitted factors included in u_t can themselves be serially correlated. For example, suppose that the demand for orange juice also depends on income, so one factor that influences the price of orange juice is income, specifically, the aggregate income of potential orange juice consumers. Then aggregate income is an omitted variable in the distributed lag regression of orange juice price changes against freezing degree days. Aggregate income, however, is serially correlated: Income tends to fall in recessions and rise in expansions. Thus, income is serially correlated, and, because it is part of the error term, u_t will be serially correlated. This example is typical: Because omitted determinants of Y are themselves serially correlated, in general u_t in the distributed lag model will be correlated.

The autocorrelation of u_t does not affect the consistency of OLS, nor does it introduce bias. If, however, the errors are autocorrelated, then in general the usual OLS standard errors are inconsistent and a different formula must be used. Thus correlation of the errors is analogous to heteroskedasticity: The homoskedasticity-only standard errors are "wrong" when the errors are in fact heteroskedastic, in the sense that using homoskedasticity-only standard errors results in misleading statistical inferences when the errors are heteroskedastic. Similarly, when the errors are serially correlated, standard errors predicated upon i.i.d. errors are "wrong" in the sense that they result in misleading statistical inferences. The solution to this problem is to use heteroskedasticity- and autocorrelation-consistent (HAC) standard errors, the topic of Section 15.4.

Dynamic Multipliers and Cumulative Dynamic Multipliers

Another name for the dynamic causal effect is the dynamic multiplier. The cumulative dynamic multipliers are the cumulative causal effects, up to a given lag; thus the cumulative dynamic multipliers measure the cumulative effect on Y of a change in X.

Dynamic multipliers. The effect of a unit change in X on Y after h periods, which is β_{h+1} in Equation (15.4), is called the h-period **dynamic multiplier**. Thus the dynamic multipliers relating X to Y are the coefficients on X_t and its lags in Equation (15.4). For example, β_2 is the one-period dynamic multiplier, β_3 is the two-period dynamic multiplier, and so forth. In this terminology, the zero-period (or contemporaneous) dynamic multiplier, or **impact effect**, is β_1, the effect on Y of a change in X in the same period.

Because the dynamic multipliers are estimated by the OLS regression coefficients, their standard errors are the HAC standard errors of the OLS regression coefficients.

Cumulative dynamic multipliers. The h-period **cumulative dynamic multiplier** is the cumulative effect of a unit change in X on Y over the next h periods. Thus the cumulative dynamic multipliers are the cumulative sum of the dynamic multipliers. In terms of the coefficients of the distributed lag regression in Equation (15.4), the zero-period cumulative multiplier is β_1, the one-period cumulative multiplier is $\beta_1 + \beta_2$, and the h-period cumulative dynamic multiplier is $\beta_1 + \beta_2 + \cdots + \beta_{h+1}$. The sum of all the individual dynamic multipliers, $\beta_1 + \beta_2 + \cdots + \beta_{r+1}$, is the cumulative long-run effect on Y of a change in X and is called the **long-run cumulative dynamic multiplier**.

For example, consider the regression in Equation (15.2). The immediate effect of an additional freezing degree day is that the price of orange juice concentrate rises by 0.47%. The cumulative effect of a price change over the next month is the sum of the impact effect and the dynamic effect one month ahead; thus the cumulative effect on prices is the initial increase of 0.47% plus the subsequent smaller increase of 0.14% for a total of 0.61%. Similarly, the cumulative dynamic multiplier over 2 months is 0.47% + 0.14% + 0.06% = 0.67%.

The cumulative dynamic multipliers can be estimated directly using a modification of the distributed lag regression in Equation (15.4). This modified regression is

$$Y_t = \delta_0 + \delta_1 \Delta X_t + \delta_2 \Delta X_{t-1} + \delta_3 \Delta X_{t-2} + \cdots + \delta_r \Delta X_{t-r+1} + \delta_{r+1} X_{t-r} + u_t. \quad (15.7)$$

The coefficients in Equation (15.7), $\delta_1, \delta_2, \ldots, \delta_{r+1}$, are in fact the cumulative dynamic multipliers. This can be shown by a bit of algebra (Exercise 15.5), which demonstrates that the population regressions in Equations (15.7) and (15.4) are equivalent, where $\delta_0 = \beta_0, \delta_1 = \beta_1, \delta_2 = \beta_1 + \beta_2, \delta_3 = \beta_1 + \beta_2 + \beta_3$, and so forth. The coefficient on X_{t-r}, δ_{r+1}, is the long-run cumulative dynamic multiplier; that is, $\delta_{r+1} = \beta_1 + \beta_2 + \beta_3 + \cdots + \beta_{r+1}$. Moreover, the OLS estimators of the coefficients in Equation (15.7) are the same as the corresponding cumulative sum of the OLS estimators in Equation (15.4). For example, $\hat{\delta}_2 = \hat{\beta}_1 + \hat{\beta}_2$. The main benefit of estimating the cumulative dynamic multipliers using the specification in Equation (15.7) is that, because the OLS estimators of the regression coefficients are estimators of the cumulative dynamic multipliers, the HAC standard errors of the coefficients in Equation (15.7) are the HAC standard errors of the cumulative dynamic multipliers.

15.4 Heteroskedasticity- and Autocorrelation-Consistent Standard Errors

If the error term u_t is autocorrelated, then OLS coefficient estimators are consistent, but in general the usual OLS standard errors for cross-sectional data are not. This means that conventional statistical inferences—hypothesis tests and confidence intervals—based on the usual OLS standard errors will, in general, be misleading. For example, confidence intervals constructed as the OLS estimator

± 1.96 conventional standard errors need not contain the true value in 95% of repeated samples, even if the sample size is large. This section begins with a derivation of the correct formula for the variance of the OLS estimator with autocorrelated errors, then turns to heteroskedasticity- and autocorrelation-consistent (HAC) standard errors.

This section covers HAC standard errors for regression with time series data. Chapter 10 introduced a type of HAC standard errors, clustered standard errors, which are appropriate for panel data. Although clustered standard errors for panel data and HAC standard errors for time series data have the same goal, the different data structures lead to different formulas. This section is self-contained, and Chapter 10 is not a prerequisite.

Distribution of the OLS Estimator with Autocorrelated Errors

To keep things simple, consider the OLS estimator $\hat{\beta}_1$ in the distributed lag regression model with no lags, that is, the linear regression model with a single regressor X_t:

$$Y_t = \beta_0 + \beta_1 X_t + u_t, \tag{15.8}$$

where the assumptions of Key Concept 15.2 are satisfied. This section shows that the variance of $\hat{\beta}_1$ can be written as the product of two terms: the expression for $\text{var}(\hat{\beta}_1)$, applicable if u_t is not serially correlated, multiplied by a correction factor that arises from the autocorrelation in u_t or, more precisely, the autocorrelation in $(X_t - \mu_X)u_t$.

As shown in Appendix 4.3, the formula for the OLS estimator $\hat{\beta}_1$ in Key Concept 4.2 can be rewritten as

$$\hat{\beta}_1 = \beta_1 + \frac{\frac{1}{T}\sum_{t=1}^{T}(X_t - \overline{X})u_t}{\frac{1}{T}\sum_{t=1}^{T}(X_t - \overline{X})^2}, \tag{15.9}$$

where Equation (15.9) is Equation (4.30) with a change of notation so that i and n are replaced by t and T. Because $\overline{X} \xrightarrow{p} \mu_X$ and $\frac{1}{T}\sum_{t=1}^{T}(X_t - \overline{X})^2 \xrightarrow{p} \sigma_X^2$, in large samples $\hat{\beta}_1 - \beta_1$ is approximately given by

$$\hat{\beta}_1 - \beta_1 \cong \frac{\frac{1}{T}\sum_{t=1}^{T}(X_t - \mu_X)u_t}{\sigma_X^2} = \frac{\frac{1}{T}\sum_{t=1}^{T}v_t}{\sigma_X^2} = \frac{\overline{v}}{\sigma_X^2}, \tag{15.10}$$

where $v_t = (X_t - \mu_X)u_t$ and $\overline{v} = \frac{1}{T}\sum_{t=1}^{T}v_t$. Thus

$$\text{var}(\hat{\beta}_1) = \text{var}\left(\frac{\overline{v}}{\sigma_X^2}\right) = \frac{\text{var}(\overline{v})}{(\sigma_X^2)^2}. \tag{15.11}$$

If v_t is i.i.d.—as assumed for cross-sectional data in Key Concept 4.3—then $\text{var}(\overline{v}) = \text{var}(v_t)/T$ and the formula for the variance of $\hat{\beta}_1$ from Key Concept 4.4 applies. If, however, u_t and X_t are not independently distributed over time, then in general v_t will be serially correlated, so $\text{var}(\overline{v}) \neq \text{var}(v_t)/T$ and Key Concept 4.4 does not apply. Instead, if v_t is serially correlated, the variance of $\overline{v}$ is given by

$$\begin{aligned}
\text{var}(\overline{v}) &= \text{var}[(v_1 + v_2 + \cdots + v_T)/T] \\
&= [\text{var}(v_1) + \text{cov}(v_1, v_2) + \cdots + \text{cov}(v_1, v_T) \\
&\quad + \text{cov}(v_2, v_1) + \text{var}(v_2) + \cdots + \text{var}(v_T)]/T^2 \\
&= [T\text{var}(v_t) + 2(T-1)\text{cov}(v_t, v_{t-1}) \\
&\quad + 2(T-2)\text{cov}(v_t, v_{t-2}) + \cdots + 2\text{cov}(v_t, v_{t-T+1})]/T^2 \\
&= \frac{\sigma_v^2}{T} f_T,
\end{aligned} \tag{15.12}$$

where

$$f_T = 1 + 2\sum_{j=1}^{T-1}\left(\frac{T-j}{T}\right)\rho_j, \tag{15.13}$$

where $\rho_j = \text{corr}(v_t, v_{t-j})$. In large samples, f_T tends to the limit, $f_T \longrightarrow f_\infty = 1 + 2\sum_{j=1}^{\infty}\rho_j$.

Combining the expressions in Equation (15.10) for $\hat{\beta}_1$ and Equation (15.12) for $\text{var}(\overline{v})$ gives the formula for the variance of $\hat{\beta}_1$ when v_t is autocorrelated:

$$\text{var}(\hat{\beta}_1) = \left[\frac{1}{T}\frac{\sigma_v^2}{(\sigma_X^2)^2}\right] f_T, \tag{15.14}$$

where f_T is given in Equation (15.13).

Equation (15.14) expresses the variance of $\hat{\beta}_1$ as the product of two terms. The first, in square brackets, is the formula for the variance of $\hat{\beta}_1$ given in Key Concept 4.4, which applies in the absence of serial correlation. The second is the factor f_T, which adjusts this formula for serial correlation. Because of this additional factor f_T in Equation (15.14), the usual OLS standard error computed using Equation (5.4) is incorrect if the errors are serially correlated: If $v_t = (X_t - \mu_X)u_t$ is serially correlated, the estimator of the variance is off by the factor f_T.

HAC Standard Errors

If the factor f_T, defined in Equation (15.13), was known, then the variance of $\hat{\beta}_1$ could be estimated by multiplying the usual cross-sectional estimator of the variance by f_T. This factor, however, depends on the unknown autocorrelations of v_t, so it must be estimated. The estimator of the variance of $\hat{\beta}_1$ that incorporates this adjustment is consistent whether or not there is heteroskedasticity and whether or not v_t is autocorrelated. Accordingly, this estimator is called the **heteroskedasticity- and autocorrelation-consistent (HAC)** estimator of the variance of $\hat{\beta}_1$, and the square root of the HAC variance estimator is the **HAC standard error** of $\hat{\beta}_1$.

The HAC variance formula. The heteroskedasticity- and autocorrelation-consistent estimator of the variance of $\hat{\beta}_1$ is

$$\widetilde{\sigma}^2_{\hat{\beta}_1} = \hat{\sigma}^2_{\hat{\beta}_1} \hat{f}_T, \tag{15.15}$$

where $\hat{\sigma}^2_{\hat{\beta}_1}$ is the estimator of the variance of $\hat{\beta}_1$ in the absence of serial correlation, given in Equation (5.4), and where $\hat{f}_T$ is an estimator of the factor f_T in Equation (15.13).

The task of constructing a consistent estimator $\hat{f}_T$ is challenging. To see why, consider two extremes. At one extreme, given the formula in Equation (15.13), it might seem natural to replace the population autocorrelations ρ_j with the sample autocorrelations $\hat{\rho}_j$ [defined in Equation (14.6)], yielding the estimator $1 + 2\sum_{j=1}^{T-1}\left(\frac{T-j}{T}\right)\hat{\rho}_j$. But this estimator contains so many estimated autocorrelations that it is inconsistent. Intuitively, because each of the estimated autocorrelations contains an estimation error, by estimating so many autocorrelations the estimation error in this estimator of f_T remains large even in large samples. At the other extreme, one could imagine using only a few sample autocorrelations, for example, only the first sample autocorrelation, and ignoring all the higher autocorrelations. Although this estimator eliminates the problem of estimating too many autocorrelations, it has a different problem: It is inconsistent because it ignores the additional autocorrelations that appear in Equation (15.13). In short, using too many sample autocorrelations makes the estimator have a large variance, but using too few autocorrelations ignores the autocorrelations at higher lags, so in either of these extreme cases the estimator is inconsistent.

Estimators of f_T used in practice strike a balance between these two extreme cases by choosing the number of autocorrelations to include in a way that depends on the sample size T. If the sample size is small, only a few autocorrelations are used, but if the sample size is large, more autocorrelations are included (but still far fewer than T). Specifically, let $\hat{f}_T$ be given by

$$\hat{f}_T = 1 + 2\sum_{j=1}^{m-1}\left(\frac{m-j}{m}\right)\widetilde{\rho}_j, \tag{15.16}$$

where $\widetilde{\rho}_j = \sum_{t=j+1}^{T}\hat{v}_t\hat{v}_{t-j}/\sum_{t=1}^{T}\hat{v}_t^2$, where $\hat{v}_t = (X_t - \overline{X})\hat{u}_t$ (as in the definition of $\hat{\sigma}^2_{\hat{\beta}_1}$). The parameter m in Equation (15.16) is called the **truncation parameter** of the HAC estimator because the sum of autocorrelations is shortened, or truncated, to include only $m - 1$ autocorrelations instead of the $T - 1$ autocorrelations appearing in the population formula in Equation (15.13).

For $\hat{f}_T$ to be consistent, m must be chosen so that it is large in large samples, although still much less than T. One guideline for choosing m in practice is to use the formula

$$m = 0.75T^{1/3}, \tag{15.17}$$

rounded to an integer. This formula, which is based on the assumption that there is a moderate amount of autocorrelation in v_t, gives a benchmark rule for determining m as a function of the number of observations in the regression.[1]

The value of the truncation parameter m resulting from Equation (15.17) can be modified using your knowledge of the series at hand. On the one hand, if there is a great deal of serial correlation in v_t, then you could increase m beyond the value from Equation (15.17). On the other hand, if v_t has little serial correlation, you could decrease m. Because of the ambiguity associated with the choice of m, it is good practice to try one or two alternative values of m for at least one specification to make sure your results are not sensitive to m.

The HAC estimator in Equation (15.15), with $\hat{f}_T$ given in Equation (15.16), is called the **Newey–West variance estimator**, after the econometricians Whitney Newey and Kenneth West, who proposed it. They showed that, when used along with a rule like that in Equation (15.17), under general assumptions this estimator is a consistent estimator of the variance of $\hat{\beta}_1$ (Newey and West, 1987). Their

[1]Equation (15.17) gives the "best" choice of m if u_t and X_t are first-order autoregressive processes with first autocorrelation coefficients 0.5, where "best" means the estimator that minimizes $E(\widetilde{\sigma}^2_{\hat{\beta}_1} - \sigma^2_{\hat{\beta}_1})^2$. Equation (15.17) is based on a more general formula derived by Andrews [1991, Equation (5.3)].

proofs (and those in Andrews, 1991) assume that v_t has more than four moments, which in turn is implied by X_t and u_t having more than eight moments, and this is the reason that the third assumption in Key Concept 15.2 is that X_t and u_t have more than eight moments.

Other HAC estimators. The Newey–West variance estimator is not the only HAC estimator. For example, the weights $(m - j)/m$ in Equation (15.16) can be replaced by different weights. If different weights are used, then the rule for choosing the truncation parameter in Equation (15.17) no longer applies and a different rule, developed for those weights, should be used instead. Discussion of HAC estimators using other weights goes beyond the scope of this book. For more information on this topic, see Hayashi (2000, Section 6.6).

Extension to multiple regression. All the issues discussed in this section generalize to the distributed lag regression model in Key Concept 15.1 with multiple lags and, more generally, to the multiple regression model with serially correlated errors. In particular, if the error term is serially correlated, then the usual OLS standard errors are an unreliable basis for inference and HAC standard errors should be used instead. If the HAC variance estimator used is the Newey–West estimator [the HAC variance estimator based on the weights $(m - j)/m$], then the truncation parameter m can be chosen according to the rule in Equation (15.17) whether there is a single regressor or multiple regressors. The formula for HAC standard errors in multiple regression is incorporated into modern regression software designed for use with time series data. Because this formula involves matrix algebra, we omit it here and instead refer the reader to Hayashi (2000, Section 6.6) for the mathematical details.

HAC standard errors are summarized in Key Concept 15.3.

15.5 Estimation of Dynamic Causal Effects with Strictly Exogenous Regressors

When X_t is strictly exogenous, two alternative estimators of dynamic causal effects are available. The first such estimator involves estimating an autoregressive distributed lag (ADL) model instead of a distributed lag model and calculating the dynamic multipliers from the estimated ADL coefficients. This method can entail estimating fewer coefficients than OLS estimation of the distributed lag model, thus potentially reducing estimation error. The second method is to estimate the coefficients of the distributed lag model, using **generalized least squares (GLS)**

HAC Standard Errors

KEY CONCEPT 15.3

The problem: The error term u_t in the distributed lag regression model in Key Concept 15.1 can be serially correlated. If so, the OLS coefficient estimators are consistent but in general the usual OLS standard errors are not, resulting in misleading hypothesis tests and confidence intervals.

The solution: Standard errors should be computed using a heteroskedasticity- and autocorrelation-consistent (HAC) estimator of the variance. The HAC estimator involves estimates of $m - 1$ autocovariances as well as the variance; in the case of a single regressor, the relevant formulas are given in Equations (15.15) and (15.16).

In practice, using HAC standard errors entails choosing the truncation parameter m. To do so, use the formula in Equation (15.17) as a benchmark, then increase or decrease m depending on whether your regressors and errors have high or low serial correlation.

instead of OLS. Although the same number of coefficients in the distributed lag model are estimated by GLS as by OLS, the GLS estimator has a smaller variance. To keep the exposition simple, these two estimation methods are initially laid out and discussed in the context of a distributed lag model with a single lag and AR(1) errors. The potential advantages of these two estimators are greatest, however, when many lags appear in the distributed lag model, so these estimators are then extended to the general distributed lag model with higher-order autoregressive errors.

The Distributed Lag Model with AR(1) Errors

Suppose that the causal effect on Y of a change in X lasts for only two periods; that is, it has an initial impact effect β_1 and an effect in the next period of β_2, but no effect thereafter. Then the appropriate distributed lag regression model is the distributed lag model with only current and past values of X_{t-1}:

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 X_{t-1} + u_t. \qquad (15.18)$$

As discussed in Section 15.2, in general the error term u_t in Equation (15.18) is serially correlated. One consequence of this serial correlation is that, if the distributed lag coefficients are estimated by OLS, then inference based on the usual

OLS standard errors can be misleading. For this reason, Sections 15.3 and 15.4 emphasized the use of HAC standard errors when β_1 and β_2 in Equation (15.18) are estimated by OLS.

In this section, we take a different approach toward the serial correlation in u_t. This approach, which is possible if X_t is strictly exogenous, involves adopting an autoregressive model for the serial correlation in u_t, then using this AR model to derive some estimators that can be more efficient than the OLS estimator in the distributed lag model.

Specifically, suppose that u_t follows the AR(1) model

$$u_t = \phi_1 u_{t-1} + \widetilde{u}_t, \tag{15.19}$$

where ϕ_1 is the autoregressive parameter, $\widetilde{u}_t$ is serially uncorrelated, and no intercept is needed because $E(u_t) = 0$. Equations (15.18) and (15.19) imply that the distributed lag model with a serially correlated error can be rewritten as an autoregressive distributed lag model with a serially uncorrelated error. To do so, lag each side of Equation (15.18) and subtract ϕ_1 multiplied by this lag from each side:

$$\begin{aligned} Y_t - \phi_1 Y_{t-1} &= (\beta_0 + \beta_1 X_t + \beta_2 X_{t-1} + u_t) - \phi_1(\beta_0 + \beta_1 X_{t-1} + \beta_2 X_{t-2} + u_{t-1}) \\ &= \beta_0 + \beta_1 X_t + \beta_2 X_{t-1} - \phi_1\beta_0 - \phi_1\beta_1 X_{t-1} - \phi_1\beta_2 X_{t-2} + \widetilde{u}_t, \end{aligned} \tag{15.20}$$

where the second equality uses $\widetilde{u}_t = u_t - \phi_1 u_{t-1}$. Collecting terms in Equation (15.20), we have that

$$Y_t = \alpha_0 + \phi_1 Y_{t-1} + \delta_0 X_t + \delta_1 X_{t-1} + \delta_2 X_{t-2} + \widetilde{u}_t, \tag{15.21}$$

where

$$\alpha_0 = \beta_0(1 - \phi_1),\ \delta_0 = \beta_1,\ \delta_1 = \beta_2 - \phi_1\beta_1, \text{ and } \delta_2 = -\phi_1\beta_2, \tag{15.22}$$

where β_0, β_1, and β_2 are the coefficients in Equation (15.18) and ϕ_1 is the autocorrelation coefficient in Equation (15.19).

Equation (15.21) is an ADL model that includes a contemporaneous value of X and two of its lags. We will refer to Equation (15.21) as the ADL representation of the distributed lag model with autoregressive errors given in Equations (15.18) and (15.19).

The terms in Equation (15.20) can be reorganized differently to obtain an expression that is equivalent to Equations (15.21) and (15.22). Let $\widetilde{Y}_t = Y_t - \phi_1 Y_{t-1}$ be the **quasi-difference** of Y_t ("quasi" because it is not the first difference, the difference between Y_t and Y_{t-1}; rather, it is the difference between Y_t and $\phi_1 Y_{t-1}$).

Similarly, let $\widetilde{X}_t = X_t - \phi_1 X_{t-1}$ be the quasi-difference of X_t. Then Equation (15.20) can be written

$$\widetilde{Y}_t = \alpha_0 + \beta_1 \widetilde{X}_t + \beta_2 \widetilde{X}_{t-1} + \widetilde{u}_t. \tag{15.23}$$

We will refer to Equation (15.23) as the quasi-difference representation of the distributed lag model with autoregressive errors given in Equations (15.18) and (15.19).

The ADL model Equation (15.21) [with the parameter restrictions in Equation (15.22)] and the quasi-difference model in Equation (15.23) are equivalent. In both models, the error term, $\widetilde{u}_t$, is serially uncorrelated. The two representations, however, suggest different estimation strategies. But before discussing those strategies, we turn to the assumptions under which they yield consistent estimators of the dynamic multipliers, β_1 and β_2.

The conditional mean zero assumption in the ADL(1,2) and quasi-differenced models. Because Equations (15.21) [with the restrictions in Equation (15.22)] and (15.23) are equivalent, the conditions for their estimation are the same, so for convenience we consider Equation (15.23).

The quasi-difference model in Equation (15.23) is a distributed lag model involving the quasi-differenced variables with a serially uncorrelated error. Accordingly, the conditions for OLS estimation of the coefficients in Equation (15.23) are the least squares assumptions for the distributed lag model in Key Concept 15.2, expressed in terms of $\widetilde{u}_t$ and $\widetilde{X}_t$. The critical assumption here is the first assumption, which, applied to Equation (15.23), is that $\widetilde{X}_t$ is exogenous; that is,

$$E(\widetilde{u}_t|\widetilde{X}_t, \widetilde{X}_{t-1}, \ldots) = 0, \tag{15.24}$$

where letting the conditional expectation depend on distant lags of $\widetilde{X}_t$ ensures that no additional lags of $\widetilde{X}_t$, other than those appearing in Equation (15.23), enter the population regression function.

Because $\widetilde{X}_t = X_t - \phi_1 X_{t-1}$, so $X_t = \widetilde{X}_t + \phi_1 X_{t-1}$, conditioning on $\widetilde{X}_t$ and all of its lags is equivalent to conditioning on X_t and all of its lags. Thus the conditional expectation condition in Equation (15.24) is equivalent to the condition that $E(\widetilde{u}_t|X_t, X_{t-1}, \ldots) = 0$. Furthermore, because $\widetilde{u}_t = u_t - \phi_1 u_{t-1}$, this condition in turn implies that

$$\begin{aligned} 0 &= E(\widetilde{u}_t|X_t, X_{t-1}, \ldots) \\ &= E(u_t - \phi_1 u_{t-1}|X_t, X_{t-1}, \ldots) \\ &= E(u_t|X_t, X_{t-1}, \ldots) - \phi_1 E(u_{t-1}|X_t, X_{t-1}, \ldots). \end{aligned} \tag{15.25}$$

For the equality in Equation (15.25) to hold for general values of ϕ_1, it must be the case that both $E(u_t|X_t, X_{t-1}, \dots) = 0$ and $E(u_{t-1}|X_t, X_{t-1}, \dots) = 0$. By shifting the time subscripts, the condition that $E(u_{t-1}|X_t, X_{t-1}, \dots) = 0$ can be rewritten as

$$E(u_t|X_{t+1}, X_t, X_{t-1}, \dots) = 0, \tag{15.26}$$

which (by the law of iterated expectations) implies that $E(u_t|X_t, X_{t-1}, \dots) = 0$. In summary, having the zero conditional mean assumption in Equation (15.24) hold for general values of ϕ_1 is equivalent to having the condition in Equation (15.26) hold.

The condition in Equation (15.26) is implied by X_t being strictly exogenous, but it is *not* implied by X_t being (past and present) exogenous. Thus the least squares assumptions for estimation of the distributed lag model in Equation (15.23) hold if X_t is strictly exogenous, but it is not enough that X_t be (past and present) exogenous.

Because the ADL representation [Equations (15.21) and (15.22)] is equivalent to the quasi-differenced representation [Equation (15.23)], the conditional mean assumption needed to estimate the coefficients of the quasi-differenced representation [that $E(u_t|X_{t+1}, X_t, X_{t-1}, \dots) = 0$] is also the conditional mean assumption for consistent estimation of the coefficients of the ADL representation.

We now turn to the two estimation strategies suggested by these two representations: estimation of the ADL coefficients and estimation of the coefficients of the quasi-differenced model.

OLS Estimation of the ADL Model

The first strategy is to use OLS to estimate the coefficients in the ADL model in Equation (15.21). As the derivation leading to Equation (15.21) shows, including the lag of Y and the extra lag of X as regressors makes the error term serially uncorrelated (under the assumption that the error follows a first order autoregression). Thus the usual OLS standard errors can be used; that is, HAC standard errors are not needed when the ADL model coefficients in Equation (15.21) are estimated by OLS.

The estimated ADL coefficients are not themselves estimates of the dynamic multipliers, but the dynamic multipliers can be computed from the ADL coefficients. A general way to compute the dynamic multipliers is to express the estimated regression function as a function of current and past values of X_t, that is, to eliminate Y_t from the estimated regression function. To do so, repeatedly substitute expressions for lagged values of Y_t into the estimated regression function. Specifically, consider the estimated regression function

$$\hat{Y}_t = \hat{\phi}_1 Y_{t-1} + \hat{\delta}_0 X_t + \hat{\delta}_1 X_{t-1} + \hat{\delta}_2 X_{t-2}, \tag{15.27}$$

where the estimated intercept has been omitted because it does not enter any expression for the dynamic multipliers. Lagging both sides of Equation (15.27) yields $\hat{Y}_{t-1} = \hat{\phi}_1 Y_{t-2} + \hat{\delta}_0 X_{t-1} + \hat{\delta}_1 X_{t-2} + \hat{\delta}_2 X_{t-3}$, so replacing $\hat{Y}_{t-1}$ in Equation (15.27) by this expression for $\hat{Y}_{t-1}$ and collecting terms yields

$$\begin{aligned}\hat{Y}_t &= \hat{\phi}_1(\hat{\phi}_1 Y_{t-2} + \hat{\delta}_0 X_{t-1} + \hat{\delta}_1 X_{t-2} + \hat{\delta}_2 X_{t-3}) + \hat{\delta}_0 X_t + \hat{\delta}_1 X_{t-1} + \hat{\delta}_2 X_{t-2} \\ &= \hat{\delta}_0 X_t + (\hat{\delta}_1 + \hat{\phi}_1\hat{\delta}_0) X_{t-1} + (\hat{\delta}_2 + \hat{\phi}_1\hat{\delta}_1) X_{t-2} + \hat{\phi}_1\hat{\delta}_2 X_{t-3} + \hat{\phi}_1^2 Y_{t-2}.\end{aligned} \tag{15.28}$$

Repeating this process by repeatedly substituting expressions for Y_{t-2}, Y_{t-3}, and so forth yields

$$\begin{aligned}\hat{Y}_t &= \hat{\delta}_0 X_t + (\hat{\delta}_1 + \hat{\phi}_1\hat{\delta}_0) X_{t-1} + (\hat{\delta}_2 + \hat{\phi}_1\hat{\delta}_1 + \hat{\phi}_1^2\hat{\delta}_0) X_{t-2} \\ &\quad + \hat{\phi}_1(\hat{\delta}_2 + \hat{\phi}_1\hat{\delta}_1 + \hat{\phi}_1^2\hat{\delta}_0) X_{t-3} + \hat{\phi}_1^2(\hat{\delta}_2 + \hat{\phi}_1\hat{\delta}_1 + \hat{\phi}_1^2\hat{\delta}_0) X_{t-4} + \cdots.\end{aligned} \tag{15.29}$$

The coefficients in Equation (15.29) are the estimators of the dynamic multipliers, computed from the OLS estimators of the coefficients in the ADL model in Equation (15.21). If the restrictions on the coefficients in Equation (15.22) were to hold exactly for the *estimated* coefficients, then the dynamic multipliers beyond the second (that is, the coefficients on X_{t-2}, X_{t-3}, and so forth) would all be zero.[2] However, under this estimation strategy those restrictions will not hold exactly, so the estimated multipliers beyond the second in Equation (15.29) will generally be nonzero.

GLS Estimation

The second strategy for estimating the dynamic multipliers when X_t is strictly exogenous is to use generalized least squares (GLS), which entails estimating Equation (15.23). To describe the GLS estimator, we initially assume that ϕ_1 is known. Because in practice it is unknown, this estimator is infeasible, so it is called the infeasible GLS estimator. The infeasible GLS estimator, however, can be modified using an estimator of ϕ_1, which yields a feasible version of the GLS estimator.

[2]Substitute the equalities in Equation (15.22) to show that, if those equalities hold, then $\delta_2 + \phi_1\delta_1 + \phi_1^2\delta_0 = 0$.

Infeasible GLS. Suppose that ϕ_1 were known; then the quasi-differenced variables $\tilde{X}_t$ and $\tilde{Y}_t$ could be computed directly. As discussed in the context of Equations (15.24) and (15.26), if X_t is strictly exogenous, then $E(\tilde{u}_t|\tilde{X}_t, \tilde{X}_{t-1}, \ldots) = 0$. Thus, if X_t is strictly exogenous and if ϕ_1 is known, the coefficients α_0, β_1, and β_2 in Equation (15.23) can be estimated by the OLS regression of $\tilde{Y}_t$ on $\tilde{X}_t$ and $\tilde{X}_{t-1}$ (including an intercept). The resulting estimator of β_1 and β_2—that is, the OLS estimator of the slope coefficients in Equation (15.23) when ϕ_1 is known—is the **infeasible GLS estimator**. This estimator is infeasible because ϕ_1 is unknown, so $\tilde{X}_t$ and $\tilde{Y}_t$ cannot be computed and thus these OLS estimators cannot actually be computed.

Feasible GLS. The **feasible GLS estimator** modifies the infeasible GLS estimator by using a preliminary estimator of ϕ_1, $\hat{\phi}_1$, to compute the estimated quasi-differences. Specifically, the feasible GLS estimators of β_1 and β_2 are the OLS estimators of β_1 and β_2 in Equation (15.23), computed by regressing $\hat{\tilde{Y}}_t$ on $\hat{\tilde{X}}_t$ and $\hat{\tilde{X}}_{t-1}$ (with an intercept), where $\hat{\tilde{X}}_t = X_t - \hat{\phi}_1 X_{t-1}$ and $\hat{\tilde{Y}}_t = Y_t - \hat{\phi}_1 Y_{t-1}$.

The preliminary estimator, $\hat{\phi}_1$, can be computed by first estimating the distributed lag regression in Equation (15.18) by OLS, then using OLS to estimate ϕ_1 in Equation (15.19) with the OLS residuals $\hat{u}_t$ replacing the unobserved regression errors u_t. This version of the GLS estimator is called the Cochrane–Orcutt (1949) estimator.

An extension of the Cochrane–Orcutt method is to continue this process iteratively: Use the GLS estimator of β_1 and β_2 to compute revised estimators of u_t; use these new residuals to re-estimate ϕ_1; use this revised estimator of ϕ_1 to compute revised estimated quasi-differences; use these revised estimated quasi-differences to re-estimate β_1 and β_2; and continue this process until the estimators of β_1 and β_2 converge. This is referred to as the iterated Cochrane–Orcutt estimator.

A nonlinear least squares interpretation of the GLS estimator. An equivalent interpretation of the GLS estimator is that it estimates the ADL model in Equation (15.21), imposing the parameter restrictions in Equation (15.22). These restrictions are nonlinear functions of the original parameters β_0, β_1, β_2, and ϕ_1, so this estimation cannot be performed using OLS. Instead, the parameters can be estimated by nonlinear least squares (NLLS). As discussed in Appendix 8.1, NLLS minimizes the sum of squared mistakes made by the estimated regression function, recognizing that the regression function is a nonlinear function of the parameters being estimated. In general, NLLS estimation can require sophisticated algorithms for minimizing nonlinear functions of unknown parameters. In the special case at hand, however, those sophisticated algorithms are not needed; rather, the NLLS estimator can be computed using the algorithm described previously for the iterated Cochrane–Orcutt estimator. Thus the iterated Cochrane–Orcutt GLS estimator is in fact the NLLS estimator of the ADL coefficients, subject to the nonlinear constraints in Equation (15.22).

Efficiency of GLS. The virtue of the GLS estimator is that when X is strictly exogenous and the transformed errors $\tilde{u}_t$ are homoskedastic, it is efficient among linear estimators, at least in large samples. To see this, first consider the infeasible GLS estimator. If $\tilde{u}_t$ is homoskedastic, if ϕ_1 is known (so that $\tilde{X}_t$ and $\tilde{Y}_t$ can be treated as if they are observed), and if X_t is strictly exogenous, then the Gauss–Markov theorem implies that the OLS estimator of α_0, β_1, and β_2 in Equation (15.23) is efficient among all linear conditionally unbiased estimators; that is, the OLS estimator of the coefficients in Equation (15.23) is the best linear unbiased estimator, or BLUE (Section 5.5). Because the OLS estimator of Equation (15.23) is the infeasible GLS estimator, this means that the infeasible GLS estimator is BLUE. The feasible GLS estimator is similar to the infeasible GLS estimator, except that ϕ_1 is estimated. Because the estimator of ϕ_1 is consistent and its variance is inversely proportional to T, the feasible and infeasible GLS estimators have the same variances in large samples. In this sense, if X is strictly exogenous, then the feasible GLS estimator is BLUE in large samples. In particular, if X is strictly exogenous, then GLS is more efficient than the OLS estimator of the distributed lag coefficients discussed in Section 15.3.

The Cochrane–Orcutt and iterated Cochrane–Orcutt estimators presented here are special cases of GLS estimation. In general, GLS estimation involves transforming the regression model so that the errors are homoskedastic and serially uncorrelated, then estimating the coefficients of the transformed regression model by OLS. In general, the GLS estimator is consistent and BLUE in large samples if X is strictly exogenous, but is not consistent if X is only (past and present) exogenous. The mathematics of GLS involve matrix algebra, so they are postponed to Section 18.6.

The Distributed Lag Model with Additional Lags and AR(p) Errors

The foregoing discussion of the distributed lag model in Equations (15.18) and (15.19), which has a single lag of X_t and an AR(1) error term, carries over to the general distributed lag model with multiple lags and an AR(p) error term.

The general distributed lag model with autoregressive errors. The general distributed lag model with r lags and an AR(p) error term is

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 X_{t-1} + \cdots + \beta_{r+1} X_{t-r} + u_t, \quad (15.30)$$

$$u_t = \phi_1 u_{t-1} + \phi_2 u_{t-2} + \cdots + \phi_p u_{t-p} + \widetilde{u}_t, \quad (15.31)$$

where $\beta_1, \ldots, \beta_{r+1}$ are the dynamic multipliers and $\phi_1, \ldots, \phi_p$ are the autoregressive coefficients of the error term. Under the AR(p) model for the errors, $\widetilde{u}_t$ is serially uncorrelated.

Algebra of the sort that led to the ADL model in Equation (15.21) shows that Equations (15.30) and (15.31) imply that Y_t can be written in ADL form:

$$Y_t = \alpha_0 + \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + \delta_0 X_t + \delta_1 X_{t-1} + \cdots + \delta_q X_{t-q} + \widetilde{u}_t, \quad (15.32)$$

where $q = r + p$ and $\delta_0, \ldots, \delta_q$ are functions of the β's and ϕ's in Equations (15.30) and (15.31). Equivalently, the model of Equations (15.30) and (15.31) can be written in quasi-difference form as

$$\widetilde{Y}_t = \alpha_0 + \beta_1 \widetilde{X}_t + \beta_2 \widetilde{X}_{t-1} + \cdots + \beta_{r+1} \widetilde{X}_{t-r} + \widetilde{u}_t, \quad (15.33)$$

where $\widetilde{Y}_t = Y_t - \phi_1 Y_{t-1} - \cdots - \phi_p Y_{t-p}$ and $\widetilde{X}_t = X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p}$.

Conditions for estimation of the ADL coefficients. The foregoing discussion of the conditions for consistent estimation of the ADL coefficients in the AR(1) case extends to the general model with AR(p) errors. The conditional mean zero assumption for Equation (15.33) is that

$$E(\widetilde{u}_t | \widetilde{X}_t, \widetilde{X}_{t-1}, \ldots) = 0. \quad (15.34)$$

Because $\widetilde{u}_t = u_t - \phi_1 u_{t-1} - \phi_2 u_{t-2} - \cdots - \phi_p u_{t-p}$ and $\widetilde{X}_t = X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p}$, this condition is equivalent to

$$E(u_t | X_t, X_{t-1}, \ldots) - \phi_1 E(u_{t-1} | X_t, X_{t-1}, \ldots) - \cdots - \phi_p E(u_{t-p} | X_t, X_{t-1}, \ldots) = 0. \quad (15.35)$$

For Equation (15.35) to hold for general values of $\phi_1, \ldots, \phi_p$, it must be the case that each of the conditional expectations in Equation (15.35) is zero; equivalently, it must be the case that

$$E(u_t | X_{t+p}, X_{t+p-1}, X_{t+p-2}, \ldots) = 0. \quad (15.36)$$

This condition is not implied by X_t being (past and present) exogenous, but it is implied by X_t being strictly exogenous. In fact, in the limit when p is infinite (so that the error term in the distributed lag model follows an infinite-order autoregression), the condition in Equation (15.36) becomes the condition in Key Concept 15.1 for strict exogeneity.

Estimation of the ADL model by OLS. As in the distributed lag model with a single lag and an AR(1) error term, the dynamic multipliers can be estimated from the OLS estimators of the ADL coefficients in Equation (15.32). The general formulas are similar to, but more complicated than, those in Equation (15.29) and are best expressed using lag multiplier notation; these formulas are given in Appendix 15.2. In practice, modern regression software designed for time series regression analysis does these computations for you.

Estimation by GLS. Alternatively, the dynamic multipliers can be estimated by (feasible) GLS. This entails OLS estimation of the coefficients of the quasi-differenced specification in Equation (15.33), using estimated quasi-differences. The estimated quasi-differences can be computed using preliminary estimators of the autoregressive coefficients $\phi_1, \ldots, \phi_p$, as in the AR(1) case. The GLS estimator is asymptotically BLUE, in the sense discussed earlier for the AR(1) case.

Estimation of dynamic multipliers under strict exogeneity is summarized in Key Concept 15.4.

Which to use: OLS or GLS? The two estimation options, OLS estimation of the ADL coefficients and GLS estimation of the distributed lag coefficients, have both advantages and disadvantages.

The advantage of the ADL approach is that it can reduce the number of parameters needed for estimating the dynamic multipliers, compared to OLS estimation of the distributed lag model. For example, the estimated ADL model in Equation (15.27) led to the infinitely long estimated distributed lag representation in Equation (15.29). To the extent that a distributed lag model with only r lags is really an approximation to a longer-lagged distributed lag model, the ADL model can provide a simple way to estimate those many longer lags using only a few unknown parameters. Thus in practice it might be possible to estimate the ADL model in Equation (15.39) with values of p and q much smaller than the value of r needed for OLS estimation of the distributed lag coefficients in Equation (15.37). In other words, the ADL specification can provide a compact, or parsimonious, summary of a long and complex distributed lag (see Appendix 15.2 for additional discussion).

KEY CONCEPT 15.4

Estimation of Dynamic Multipliers Under Strict Exogeneity

The general distributed lag model with r lags and AR(p) error term is

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 X_{t-1} + \cdots + \beta_{r+1} X_{t-r} + u_t \quad (15.37)$$

$$u_t = \phi_1 u_{t-1} + \phi_2 u_{t-2} + \cdots + \phi_p u_{t-p} + \tilde{u}_t. \quad (15.38)$$

If X_t is strictly exogenous, then the dynamic multipliers $\beta_1, \ldots, \beta_{r+1}$ can be estimated by first using OLS to estimate the coefficients of the ADL model

$$Y_t = \alpha_0 + \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + \delta_0 X_t + \delta_1 X_{t-1} + \cdots + \delta_q X_{t-q} + \tilde{u}_t, \quad (15.39)$$

where $q = r + p$ and then computing the dynamic multipliers using regression software. Alternatively, the dynamic multipliers can be estimated by estimating the distributed lag coefficients in Equation (15.37) by GLS.

The advantage of the GLS estimator is that, for a given lag length r in the distributed lag model, the GLS estimator of the distributed lag coefficients is more efficient than the OLS estimator, at least in large samples. In practice, then, the advantage of using the ADL approach arises because the ADL specification can permit estimating fewer parameters than are estimated by GLS.

15.6 Orange Juice Prices and Cold Weather

This section uses the tools of time series regression to squeeze additional insights from our data on Florida temperatures and orange juice prices. First, how long lasting is the effect of a freeze on the price? Second, has this dynamic effect been stable or has it changed over the 51 years spanned by the data and, if so, how?

We begin this analysis by estimating the dynamic causal effects using the method of Section 15.3, that is, by OLS estimation of the coefficients of a distributed lag regression of the percentage change in prices ($\%ChgP_t$) on the number of freezing degree days in that month (FDD_t) and its lagged values. For the distributed lag estimator to be consistent, FDD must be (past and present) exogenous. As discussed in Section 15.2, this assumption is reasonable here. Humans cannot influence the weather, so treating the weather as if it were randomly assigned experimentally is appropriate. Because FDD is exogenous, we can estimate the dynamic causal effects by OLS estimation of the coefficients in the distributed lag model of Equation (15.4) in Key Concept 15.1.

As discussed in Sections 15.3 and 15.4, the error term can be serially correlated in distributed lag regressions, so it is important to use HAC standard errors, which adjust for this serial correlation. For the initial results, the truncation parameter for the Newey–West standard errors (m in the notation of Section 15.4) was chosen using the rule in Equation (15.17): Because there are 612 monthly observations, according to that rule $m = 0.75T^{1/3} = 0.75 \times 612^{1/3} = 6.37$, but because m must be an integer, this was rounded up to $m = 7$; the sensitivity of the standard errors to this choice of truncation parameter is investigated below.

The results of OLS estimation of the distributed lag regression of $\%ChgP_t$ on $FDD_t, FDD_{t-1}, \ldots, FDD_{t-18}$ are summarized in column (1) of Table 15.1. The coefficients of this regression (only some of which are reported in the table) are estimates of the dynamic causal effect on orange juice price changes (in percent) for the first 18 months following a unit increase in the number of freezing degree days in a month. For example, a single freezing degree day is estimated to increase prices by 0.50% over the month in which the freezing degree day occurs. The subsequent effect on price in later months of a freezing degree day is less: After 1 month the estimated effect is to increase the price by a further 0.17%, and after 2 months the estimated effect is to increase the price by an additional 0.07%. The R^2 from this regression is 0.12, indicating that much of the monthly variation in orange juice prices is not explained by current and past values of FDD.

Plots of dynamic multipliers can convey information more effectively than tables such as Table 15.1. The dynamic multipliers from column (1) of Table 15.1 are plotted in Figure 15.2a along with their 95% confidence intervals, computed as the estimated coefficient ± 1.96 HAC standard errors. After the initial sharp price rise, subsequent price rises are less, although prices are estimated to rise slightly in each of the first 6 months after the freeze. As can be seen from Figure 15.2a, for months other than the first the dynamic multipliers are not statistically significantly different from zero at the 5% significance level, although they are estimated to be positive through the seventh month.

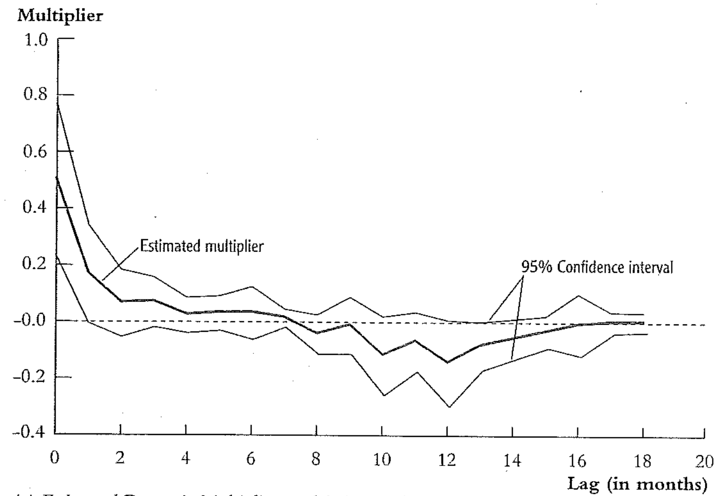
Column (2) of Table 15.1 contains the cumulative dynamic multipliers for this specification, that is, the cumulative sum of the dynamic multipliers reported in

TABLE 15.1 The Dynamic Effect of a Freezing Degree Day (*FDD*) on the Price of Orange Juice: Selected Estimated Dynamic Multipliers and Cumulative Dynamic Multipliers

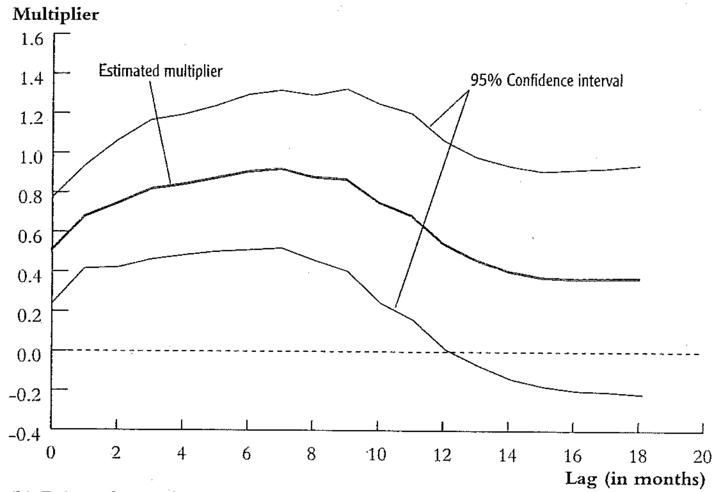
Lag Number	(1) Dynamic Multipliers	(2) Cumulative Multipliers	(3) Cumulative Multipliers	(4) Cumulative Multipliers
0	0.50 (0.14)	0.50 (0.14)	0.50 (0.14)	0.51 (0.15)
1	0.17 (0.09)	0.67 (0.14)	0.67 (0.13)	0.70 (0.15)
2	0.07 (0.06)	0.74 (0.17)	0.74 (0.16)	0.76 (0.18)
3	0.07 (0.04)	0.81 (0.18)	0.81 (0.18)	0.84 (0.19)
4	0.02 (0.03)	0.84 (0.19)	0.84 (0.19)	0.87 (0.20)
5	0.03 (0.03)	0.87 (0.19)	0.87 (0.19)	0.89 (0.20)
6 . . .	0.03 (0.05)	0.90 (0.20)	0.90 (0.21)	0.91 (0.21)
12 . .	−0.14 (0.08)	0.54 (0.27)	0.54 (0.28)	0.54 (0.28)
18	0.00 (0.02)	0.37 (0.30)	0.37 (0.31)	0.37 (0.30)
Monthly indicators?	No	No	No	Yes $F = 1.01$ $(p = 0.43)$
HAC standard error truncation parameter (m)	7	7	14	7

All regressions were estimated by OLS using monthly data (described in Appendix 15.1) from January 1950 to December 2000, for a total of $T = 612$ monthly observations. The dependent variable is the monthly percentage change in the price of orange juice ($\%ChgP_t$). Regression (1) is the distributed lag regression with the monthly number of freezing degree days and 18 of its lagged values, that is, $FDD_t, FDD_{t-1}, \ldots, FDD_{t-18}$, and the reported coefficients are the OLS estimates of the dynamic multipliers. The cumulative multipliers are the cumulative sum of estimated dynamic multipliers. All regressions include an intercept, which is not reported. Newey–West HAC standard errors, computed using the truncation number given in the final row, are reported in parentheses.

FIGURE 15.2 The Dynamic Effect of a Freezing Degree Day (*FDD*) on the Price of Orange Juice



(a) Estimated Dynamic Multipliers and 95% Confidence Interval



(b) Estimated Cumulative Dynamic Multipliers and 95% Confidence Interval

The estimated dynamic multipliers show that a freeze leads to an immediate increase in prices. Future price rises are much smaller than the initial impact. The cumulative multiplier shows that freezes have a persistent effect on the level of orange juice prices, with prices peaking seven months after the freeze.

column (1). These dynamic multipliers are plotted in Figure 15.2b along with their 95% confidence intervals. After 1 month, the cumulative effect of the freezing degree day is to increase prices by 0.67%, after 2 months the price is estimated to have risen by 0.74%, and after 6 months the price is estimated to have risen by 0.90%. As can be seen in Figure 15.2b, these cumulative multipliers increase through the seventh month, because the individual dynamic multipliers are positive for the first 7 months. In the eighth month, the dynamic multiplier is negative, so the price of orange juice begins to fall slowly from its peak. After 18 months, the cumulative increase in prices is only 0.37%; that is, the long-run cumulative dynamic multiplier is only 0.37%. This long-run cumulative dynamic multiplier is not statistically significantly different from zero at the 10% significance level ($t = 0.37/0.30 = 1.23$).

Sensitivity analysis. As in any empirical analysis, it is important to check whether these results are sensitive to changes in the details of the empirical analysis. We therefore examine three aspects of this analysis: sensitivity to the computation of the HAC standard errors; an alternative specification that investigates potential omitted variable bias; and an analysis of the stability over time of the estimated multipliers.

First, we investigate whether the standard errors reported in the second column of Table 15.1 are sensitive to different choices of the HAC truncation parameter m. In column (3), results are reported for $m = 14$, twice the value used in column (2). The regression specification is the same as in column (2), so the estimated coefficients and dynamic multipliers are identical; only the standard errors differ but, as it happens, not by much. We conclude that the results are insensitive to changes in the HAC truncation parameter.

Second, we investigate a possible source of omitted variable bias. Freezes in Florida are not randomly assigned throughout the year, but rather occur in the winter (of course). If demand for orange juice is seasonal (is demand for orange juice greater in the winter than the summer?), then the seasonal patterns in orange juice demand could be correlated with *FDD*, resulting in omitted variable bias. The quantity of oranges sold for juice is endogenous: Prices and quantities are simultaneously determined by the forces of supply and demand. Thus, as discussed in Section 9.2, including quantity would lead to simultaneity bias. Nevertheless, the seasonal component of demand can be captured by including seasonal variables as regressors. The specification in column (4) of Table 15.1 therefore includes 11 monthly binary variables, one indicating whether the month is January, one indicating February, and so forth (as usual one binary variable must be omitted to prevent perfect multicollinearity with the intercept). These monthly indicator variables are not jointly statistically significant at the 10% level ($p = 0.43$), and the estimated cumulative dynamic multipliers are essentially the same as for the specifications excluding the monthly indicators. In summary, seasonal fluctuations in demand are not an important source of omitted variable bias.

Have the dynamic multipliers been stable over time?[3] To assess the stability of the dynamic multipliers, we need to check whether the distributed lag regression coefficients have been stable over time. Because we do not have a specific break date in mind, we test for instability in the regression coefficients using the Quandt likelihood ratio (QLR) statistic (Key Concept 14.9). The QLR statistic (with 15% trimming and HAC variance estimator), computed for the regression of column (1) with all coefficients interacted, has a value of 21.19, with $q = 20$ degrees of freedom (the coefficients on FDD_t, its 18 lags, and the intercept). The 1% critical value in Table 14.6 is 2.43, so the QLR statistic rejects at the 1% significance level. These QLR regressions have 40 regressors, a large number; recomputing them for six lags only (so that there are 16 regressors and $q = 8$) also results in rejection at the 1% level. Thus the hypothesis that the dynamic multipliers are stable is rejected at the 1% significance level.

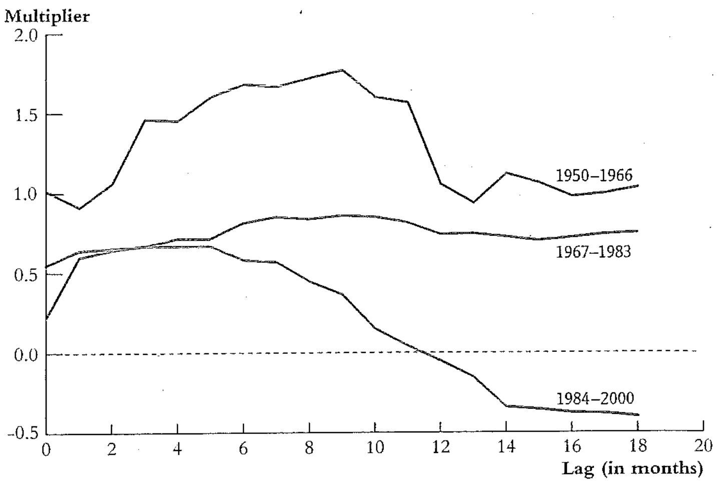
One way to see how the dynamic multipliers have changed over time is to compute them for different parts of the sample. Figure 15.3 plots the estimated cumulative dynamic multipliers for the first third (1950–1966), middle third (1967–1983), and final third (1984–2000) of the sample, computed by running separate regressions on each subsample. These estimates show an interesting and noticeable pattern. In the 1950s and early 1960s, a freezing degree day had a large and persistent effect on the price. The magnitude of the effect on price of a freezing degree day diminished in the 1970s, although it remained highly persistent. In the late 1980s and 1990s, the short-run effect of a freezing degree day was the same as in the 1970s, but it became much less persistent and was essentially eliminated after a year. These estimates suggest that the dynamic causal effect on orange juice prices of a Florida freeze became smaller and less persistent over the second half of the twentieth century. The box "Orange Trees on the March" discusses one possible explanation for the instability of the dynamic causal effects.

ADL and GLS estimates. As discussed in Section 15.5, if the error term in the distributed lag regression is serially correlated and *FDD* is strictly exogenous, it is possible to estimate the dynamic multipliers more efficiently than by OLS

[3]The discussion of stability in this subsection draws on material from Section 14.7 and can be skipped if that material has not been covered.

FIGURE 15.3 Estimated Cumulative Dynamic Multipliers from Different Sample Periods

The dynamic effect on orange juice prices of freezes changed significantly over the second half of the twentieth century. A freeze had a larger impact on prices during 1950–1966 than later, and the effect of a freeze was less persistent during 1984–2000 than earlier.



estimation of the distributed lag coefficients. Before using either the GLS estimator or the estimator based on the ADL model, however, we need to consider whether *FDD* is in fact strictly exogenous. True, humans cannot affect the daily weather, but does that mean that the weather is *strictly* exogenous? Does the error term u_t in the distributed lag regression have conditional mean zero, given past, present, and *future* values of *FDD*?

The error term in the population counterpart of the distributed lag regression in column (1) of Table 15.1 is the discrepancy between the price and its population prediction based on the past 18 months of weather. This discrepancy might arise for many reasons, one of which is that traders use forecasts of the weather in Orlando. For example, if an especially cold winter is forecasted, then traders would incorporate this into the price, so the price would be above its predicted value based on the population regression; that is, the error term would be positive. If this forecast is accurate, then in fact future weather would turn out to be cold. Thus future freezing degree days would be positive ($X_{t+1} > 0$) when the current price is unusually high ($u_t > 0$), so $\text{corr}(X_{t+1}, u_t)$ is positive. Stated more simply, although orange juice traders cannot influence the weather, they can—and do—predict it (see the box). Consequently, the error term in the price/weather regression

Orange Trees on the March

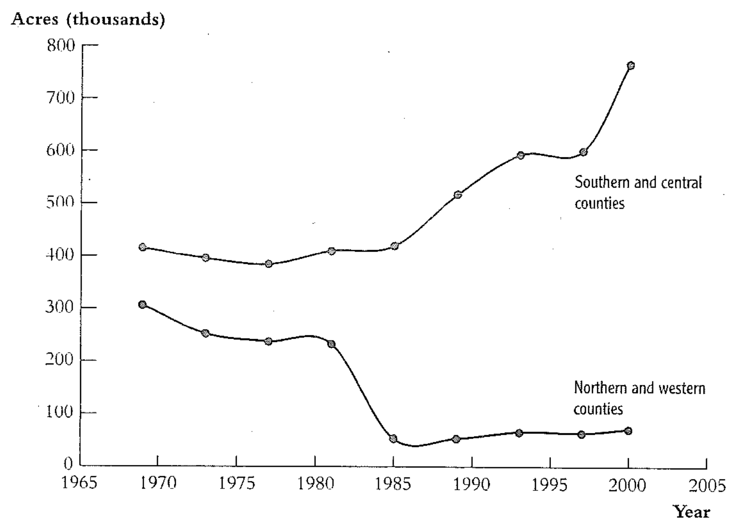
Why do the dynamic multipliers in Figure 15.3 vary over time? One possible explanation is changes in markets, but another is that the trees moved south.

According to the Florida Department of Citrus, the severe freezes in the 1980s, which are visible in Figure 15.1(c), spurred citrus growers to seek a warmer climate. As shown in Figure 15.4, the number of acres of orange trees in the more frost-prone northern and western counties fell from 232,000 acres in 1981 to 53,000 acres in 1985, and orange acreage in southern and central counties subsequently increased from 413,000 in 1985 to 588,000 in 1993. With the groves farther south, northern frosts damage a smaller fraction of the crop, and—as indicated by the dynamic multipliers in Figure 15.3—price becomes less sensitive to temperatures in the more northern city of Orlando.

OK, the orange trees themselves might not have been on the march—that can be left to *MacBeth*—but southern migration of the orange groves does give new meaning to the term "nonstationarity."[1]

[1]We are grateful to Professor James Cobbe of Florida State University for telling us about the southern movement of the orange groves.

FIGURE 15.4 Orange Grove Acreage in Regions of Florida



is correlated with future weather. In other words, *FDD* is exogenous, but if this reasoning is true, it is not strictly exogenous, and the GLS and ADL estimators will not be consistent estimators of the dynamic multipliers. These estimators therefore are not used in this application.

15.7 Is Exogeneity Plausible? Some Examples

As in regression with cross-sectional data, the interpretation of the coefficients in a distributed lag regression as causal dynamic effects hinges on the assumption that X is exogenous. If X_t or its lagged values are correlated with u_t, then the conditional mean of u_t will depend on X_t or its lags, in which case X is not (past and present) exogenous. Regressors can be correlated with the error term for several reasons, but with economic time series data a particularly important concern is that there could be simultaneous causality, which (as discussed in Sections 9.2 and 12.1) results in endogenous regressors. In Section 15.6, we discussed the assumptions of exogeneity and strict exogeneity of freezing degree days in detail. In this section, we examine the assumption of exogeneity in four other economic applications.

U.S. Income and Australian Exports

The United States is an important source of demand for Australian exports. Precisely how sensitive Australian exports are to fluctuations in U.S. aggregate income could be investigated by regressing Australian exports to the United States against a measure of U.S. income. Strictly speaking, because the world economy is integrated, there is simultaneous causality in this relationship: A decline in Australian exports reduces Australian income, which reduces demand for imports from the United States, which reduces U.S. income. As a practical matter, however, this effect is very small because the Australian economy is much smaller than the U.S. economy. Thus U.S. income plausibly can be treated as exogenous in this regression.

In contrast, in a regression of European Union exports to the United States against U.S. income, the argument for treating U.S. income as exogenous is less convincing because demand by residents of the European Union for U.S. exports constitutes a substantial fraction of the total demand for U.S. exports. Thus a decline in U.S. demand for EU exports would decrease EU income, which in turn would decrease demand for U.S. exports and thus decrease U.S. income. Because of these linkages through international trade, EU exports to the United States and U.S. income are simultaneously determined, so in this regression U.S. income arguably is not exogenous. This example illustrates a more general point that whether a variable is exogenous depends on the context: U.S. income is plausibly exogenous in a regression explaining Australian exports, but not in a regression explaining EU exports.

NEWS FLASH: Commodity Traders Send Shivers Through Disney World

Although the weather at Disney World in Orlando, Florida, is usually pleasant, now and then a cold spell can settle in. If you are visiting Disney World on a winter evening, should you bring a warm coat? Some people might check the weather forecast on TV, but those in the know can do better: They can check that day's closing price on the New York orange juice futures market!

The financial economist Richard Roll undertook a detailed study of the relationship between orange juice prices and the weather. Roll (1984) examined the effect on prices of cold weather in Orlando, but he also studied the "effect" of changes in the price of an orange juice futures contract (a contract to buy frozen orange juice concentrate at a specified date in the future) on the weather. Roll used daily data from 1975 to 1981 on the prices of OJ futures contracts traded at the New York Cotton Exchange and on daily and overnight temperatures in Orlando. He found that a rise in the price of the futures contract during the trading day in New York predicted cold weather, in particular a freezing spell, in Orlando over the following night. In fact, the market was so effective in predicting cold weather in Florida that a price rise during the trading day actually predicted forecast errors in the official U.S. government weather forecasts for that night.

Roll's study is also interesting for what he did *not* find: Although his detailed weather data explained some of the variation in daily OJ futures prices, most of the daily movements in OJ prices remained unexplained. He therefore suggested that the OJ futures market exhibits "excess volatility," that is, more volatility than can be attributed to movements in fundamentals. Understanding why (and if) there is excess volatility in financial markets is now an important area of research in financial economics.

Roll's finding also illustrates the difference between forecasting and estimating dynamic causal effects. Price changes on the OJ futures market are a useful predictor of cold weather, but that does not mean that commodity traders are so powerful that they can *cause* the temperature to fall. Visitors to Disney World might shiver after an OJ futures contract price rise, but they are not shivering *because* of the price rise—unless, of course, they went short in the OJ futures market.

Oil Prices and Inflation

Ever since the oil price increases of the 1970s, macroeconomists have been interested in estimating the dynamic effect of an increase in the international price of crude oil on the U.S. rate of inflation. Because oil prices are set in world markets in large part by foreign oil-producing countries, initially one might think that oil prices are exogenous. But oil prices are not like the weather: Members of OPEC set oil production levels strategically, taking many factors, including the state of the world economy, into account. To the extent that oil prices (or quantities) are set based on an assessment of current and future world economic conditions, including inflation in the United States, oil prices are endogenous.

Monetary Policy and Inflation

The central bankers in charge of monetary policy need to know the effect on inflation of monetary policy. Because the main tool of monetary policy is the short-term interest rate (the "short rate"), they need to know the dynamic causal effect on inflation of a change in the short rate. Although the short rate is determined by the central bank, it is not set by the central bankers at random (as it would be in an ideal randomized experiment) but rather is set endogenously: The central bank determines the short rate based on an assessment of the current and future states of the economy, especially including the current and future rates of inflation. The rate of inflation in turn depends on the interest rate (higher interest rates reduce aggregate demand), but the interest rate depends on the rate of inflation, its past value, and its (expected) future value. Thus the short rate is endogenous, and the causal dynamic effect of a change in the short rate on future inflation cannot be consistently estimated by an OLS regression of the rate of inflation on current and past interest rates.

The Phillips Curve

The Phillips curve investigated in Chapter 14 is a regression of the change in the rate of inflation against lagged changes in inflation and lags of the unemployment rate. Because lags of the unemployment rate happened in the past, one might initially think that there cannot be feedback from current rates of inflation to past values of the unemployment rate, so past values of the unemployment rate can be treated as exogenous. But past values of the unemployment rate were not randomly assigned in an experiment; instead, the past unemployment rate was simultaneously determined with past values of inflation. Because inflation and the unemployment rate are simultaneously determined, the other factors that determine inflation contained in u_t are correlated with past values of the unemployment rate; that is, the unemployment rate is not exogenous. It follows that the unemployment rate is not strictly exogenous, so the dynamic multipliers computed using an empirical Phillips curve [for example, the ADL model in Equation (14.17)] are not consistent estimates of the dynamic causal effect on inflation of a change in the unemployment rate.

15.8 Conclusion

Time series data provide the opportunity to estimate the time path of the effect on Y of a change in X, that is, the dynamic causal effect on Y of a change in X. To estimate dynamic causal effects using a distributed lag regression, however, X must be exogenous, as it would be if it were set randomly in an ideal randomized experiment. If X is not just exogenous but is *strictly* exogenous, then the dynamic causal effects can be estimated using an autoregressive distributed lag model or by GLS.

In some applications, such as estimating the dynamic causal effect on the price of orange juice of freezing weather in Florida, a convincing case can be made that the regressor (freezing degree days) is exogenous; thus the dynamic causal effect can be estimated by OLS estimation of the distributed lag coefficients. Even in this application, however, economic theory suggests that the weather is not strictly exogenous, so the ADL or GLS methods are inappropriate. Moreover, in many relations of interest to econometricians, there is simultaneous causality, so the regressor in these specifications are not exogenous, strictly or otherwise. Ascertaining whether the regressor is exogenous (or strictly exogenous) ultimately requires combining economic theory, institutional knowledge, and careful judgment.

Summary

1. Dynamic causal effects in time series are defined in the context of a randomized experiment, where the same subject (entity) receives different randomly assigned treatments at different times. The coefficients in a distributed lag regression of Y on X and its lags can be interpreted as the dynamic causal effects when the time path of X is determined randomly and independently of other factors that influence Y.
2. The variable X is (past and present) exogenous if the conditional mean of the error u_t in the distributed lag regression of Y on current and past values of X does not depend on current and past values of X. If in addition the conditional mean of u_t does not depend on future values of X, then X is strictly exogenous.
3. If X is exogenous, then the OLS estimators of the coefficients in a distributed lag regression of Y on current and past values of X are consistent estimators of the dynamic causal effects. In general, the error u_t in this regression is serially correlated, so conventional standard errors are misleading and HAC standard errors must be used instead.

4. If X is strictly exogenous, then the dynamic multipliers can be estimated by OLS estimation of an ADL model or by GLS.
5. Exogeneity is a strong assumption that often fails to hold in economic time series data because of simultaneous causality, and the assumption of strict exogeneity is even stronger.

Key Terms

dynamic causal effect (583)
distributed lag model (589)
exogeneity (590)
strict exogeneity (590)
dynamic multiplier (594)
impact effect (594)
cumulative dynamic multiplier (594)
long-run cumulative dynamic multiplier (595)
heteroskedasticity- and autocorrelation-consistent (HAC) standard error (598)
truncation parameter (599)
Newey–West variance estimator (599)
generalized least squares (GLS) (600)
quasi-difference (602)
infeasible GLS estimator (606)
feasible GLS estimator (606)

Review the Concepts

15.1 In the 1970s a common practice was to estimate a distributed lag model relating changes in nominal gross domestic product (Y) to current and past changes in the money supply (X). Under what assumptions will this regression estimate the causal effects of money on nominal GDP? Are these assumptions likely to be satisfied in a modern economy like that of the United States?

15.2 Suppose that X is strictly exogenous. A researcher estimates an ADL(1,1) model, calculates the regression residual, and finds the residual to be highly serially correlated. Should the researcher estimate a new ADL model with additional lags or simply use HAC standard errors for the ADL(1,1) estimated coefficients?

15.3 Suppose that a distributed lag regression is estimated, where the dependent variable is ΔY_t instead of Y_t. Explain how you would compute the dynamic multipliers of X_t on Y_t.

15.4 Suppose that you added FDD_{t+1} as an additional regressor in Equation (15.2). If FDD is strictly exogenous, would you expect the coefficient on FDD_{t+1} to be zero or nonzero? Would your answer change if FDD is exogenous but not strictly exogenous?

Exercises

15.1 Increases in oil prices have been blamed for several recessions in developed countries. To quantify the effect of oil prices on real economic activity, researchers have done regressions like those discussed in this chapter. Let GDP_t denote the value of quarterly gross domestic product in the United States and let $Y_t = 100\ln(GDP_t/GDP_{t-1})$ be the quarterly percentage change in GDP. James Hamilton, an econometrician and macroeconomist, has suggested that oil prices adversely affect that economy only when they jump above their values in the recent past. Specifically, let O_t equal the greater of zero or the percentage point difference between oil prices at date t and their maximum value during the past year. A distributed lag regression relating Y_t and O_t, estimated over 1955:I–2000:IV, is

$$\hat{Y}_t = \underset{(0.1)}{1.0} - \underset{(0.054)}{0.055}O_t - \underset{(0.057)}{0.026}O_{t-1} - \underset{(0.048)}{0.031}O_{t-2} - \underset{(0.042)}{0.109}O_{t-3} - \underset{(0.053)}{0.128}O_{t-4}$$
$$+ \underset{(0.025)}{0.008}O_{t-5} + \underset{(0.048)}{0.025}O_{t-6} - \underset{(0.039)}{0.019}O_{t-7} + \underset{(0.042)}{0.067}O_{t-8}.$$

a. Suppose that oil prices jump 25% above their previous peak value and stay at this new higher level (so that $O_t = 25$ and $O_{t+1} = O_{t+2} = \cdots = 0$). What is the predicted effect on output growth for each quarter over the next 2 years?
b. Construct a 95% confidence interval for your answers in (a).
c. What is the predicted cumulative change in GDP growth over eight quarters?
d. The HAC F-statistic testing whether the coefficients on O_t and its lags are zero is 3.49. Are the coefficients significantly different from zero?

15.2 Macroeconomists have also noticed that interest rates change following oil price jumps. Let R_t denote the interest rate on 3-month Treasury bills (in percentage points at an annual rate). The distributed lag regression relating the change in R_t (ΔR_t) to O_t estimated over 1955:I–2000:IV is

$$\widehat{\Delta R_t} = \underset{(0.06)}{0.07} + \underset{(0.045)}{0.062}O_t + \underset{(0.034)}{0.048}O_{t-1} - \underset{(0.028)}{0.014}O_{t-2} - \underset{(0.169)}{0.086}O_{t-3} - \underset{(0.058)}{0.000}O_{t-4}$$
$$+ \underset{(0.065)}{0.023}O_{t-5} - \underset{(0.047)}{0.010}O_{t-6} - \underset{(0.038)}{0.100}O_{t-7} - \underset{(0.025)}{0.014}O_{t-8}.$$

a. Suppose that oil prices jump 25% above their previous peak value and stay at this new higher level (so that $O_t = 25$ and $O_{t+1} = O_{t+2} = \cdots = 0$). What is the predicted change in interest rates for each quarter over the next 2 years?

b. Construct 95% confidence intervals for your answers to (a).

c. What is the effect of this change in oil prices on the level of interest rates in period $t + 8$? How is your answer related to the cumulative multiplier?

d. The HAC F-statistic testing whether the coefficients on O_t and its lags are zero is 4.25. Are the coefficients significantly different from zero?

15.3 Consider two different randomized experiments. In experiment A, oil prices are set randomly and the central bank reacts according to its usual policy rules in response to economic conditions, including changes in the oil price. In experiment B, oil prices are set randomly and the central bank holds interest rates constant and in particular does not respond to the oil price changes. In both, GDP growth is observed. Now suppose that oil prices are exogenous in the regression in Exercise 15.1. To which experiment, A or B, does the dynamic causal effect estimated in Exercise 15.1 correspond?

15.4 Suppose that oil prices are strictly exogenous. Discuss how you could improve on the estimates of the dynamic multipliers in Exercise 15.1.

15.5 Derive Equation (15.7) from Equation (15.4) and show that $\delta_0 = \beta_0$, $\delta_1 = \beta_1$, $\delta_2 = \beta_1 + \beta_2$, $\delta_3 = \beta_1 + \beta_2 + \beta_3$ (etc.). (*Hint:* Note that $X_t = \Delta X_t + \Delta X_{t-1} + \cdots + \Delta X_{t-p+1} + X_{t-p}$.)

15.6 Consider the regression model $Y_t = \beta_0 + \beta_1 X_t + u_t$, where u_t follows the stationary AR(1) model $u_t = \phi_1 u_{t-1} + \tilde{u}_t$ with $\tilde{u}_t$ i.i.d. with mean 0 and variance $\sigma^2_{\tilde{u}}$ and $|\phi_1| < 1$, the regressor X_t follows the stationary AR(1) model $X_t = \gamma_1 X_{t-1} + e_t$ with e_t i.i.d. with mean 0 and variance σ^2_e and $|\gamma_1| < 1$, and e_t is independent of $\tilde{u}_i$ for all t and i.

a. Show that $\text{var}(u_t) = \dfrac{\sigma^2_{\tilde{u}}}{1-\phi_1^2}$ and $\text{var}(X_t) = \dfrac{\sigma^2_e}{1-\gamma_1^2}$.

b. Show that $\text{cov}(u_t, u_{t-j}) = \phi_1^j \text{var}(u_t)$ and $\text{cov}(X_t, X_{t-j}) = \gamma_1^j \text{var}(X_t)$.

c. Show that $\text{corr}(u_t, u_{t-j}) = \phi_1^j$ and $\text{corr}(X_t, X_{t-j}) = \gamma_1^j$.

d. Consider the terms σ^2_v and f_T in Equation (15.14).

i. Show that $\sigma^2_v = \sigma^2_X \sigma^2_u$, where σ^2_X is the variance of X and σ^2_u is the variance of u.

ii. Derive an expression for f_∞.

15.7 Consider the regression model $Y_t = \beta_0 + \beta_1 X_t + u_t$, where u_t follows the stationary AR(1) model $u_t = \phi_1 u_{t-1} + \tilde{u}_t$ with $\tilde{u}_t$ i.i.d. with mean 0 and variance $\sigma^2_{\tilde{u}}$ and $|\phi_1| < 1$.

a. Suppose that X_t is independent of $\tilde{u}_j$ for all t and j. Is X_t exogenous (past and present)? Is X_t strictly exogenous (past, present, and future)?

b. Suppose that $X_t = \tilde{u}_{t+1}$. Is X_t exogenous? Is X_t strictly exogenous?

15.8 Consider the model in Exercise 15.7 with $X_t = \tilde{u}_{t+1}$.

a. Is the OLS estimator of β_1 consistent? Explain.

b. Explain why the GLS estimator of β_1 is not consistent.

c. Show that the infeasible GLS estimator $\hat{\beta}_1^{GLS} \xrightarrow{p} \beta_1 - \dfrac{\phi_1}{1+\phi_1^2}$. [*Hint:* Use the omitted variable formula (6.1) applied to the quasi-differenced regression Equation (15.23)].

15.9 Consider the "constant-term-only" regression model $Y_t = \beta_0 + u_t$, where u_t follows the stationary AR(1) model $u_t = \phi_1 u_{t-1} + \tilde{u}_t$ with $\tilde{u}_t$ i.i.d. with mean 0 and variance $\sigma^2_{\tilde{u}}$ and $|\phi_1| < 1$.

a. Show that the OLS estimator is $\hat{\beta}_0 = T^{-1}\sum_{t=1}^{T} Y_t$.

b. Show that the (infeasible) GLS estimator is $\hat{\beta}_0^{GLS} = (1-\phi_1)^{-1}(T-1)^{-1}\sum_{t=2}^{T-1}(Y_t - \phi_1 Y_{t-1})$. [*Hint:* The GLS estimator of β_0 is $(1-\phi_1)^{-1}$ multiplied by the OLS estimator of α_0 in Equation (15.23). Why?]

c. Show that $\hat{\beta}_0^{GLS}$ can be written as $\hat{\beta}_0^{GLS} = (T-1)^{-1}\sum_{t=2}^{T-1} Y_t + (1-\phi_1)^{-1}(T-1)^{-1}(Y_T - \phi_1 Y_1)$. [*Hint:* Rearrange the formula in (b).]

d. Derive the difference $\hat{\beta}_0 - \hat{\beta}_0^{GLS}$ and discuss why it is likely to be small when T is large.

15.10 Consider the ADL model $Y_t = 3.1 + 0.4Y_{t-1} + 2.0X_t - 0.8X_{t-1} + \tilde{u}_t$, where X_t is strictly exogenous.

a. Derive the impact effect of X on Y.

b. Derive the first five dynamic multipliers.

c. Derive the first five cumulative multipliers.

d. Derive the long-run cumulative dynamic multiplier.

Empirical Exercises

E15.1 In this exercise you will estimate the effect of oil prices on macroeconomic activity using monthly data on the Index of Industrial Production (IP) and the monthly measure of O_t described in Exercise 15.1. The data can be found on the textbook Web site **www.pearsonhighered.com/stock_watson** in the file **USMacro_Monthly**.

a. Compute the monthly growth rate in IP expressed in percentage points, $ip\_growth_t = 100 \times \ln(IP_t/IP_{t-1})$. What are the mean and standard deviation of *ip_growth* over the 1952:1–2009:12 sample period?

b. Plot the value of O_t. Why are so many values of O_t equal to zero? Why aren't some values of O_t negative?

c. Estimate a distributed lag model of *ip_growth* onto current and 18 lagged values of O_t. What value of the HAC standard truncation parameter m did you choose? Why?

d. Taken as a group, are the coefficients on O_t statistically significantly different from zero?

e. Construct graphs like those in Figure 15.2 showing the estimated dynamic multipliers, cumulative multipliers, and 95% confidence intervals. Comment on the real-world size of the multipliers.

f. Suppose that high demand in the United States (evidenced by large values of *ip_growth*) leads to increases in oil prices. Is O_t exogenous? Are the estimated multipliers shown in the graphs in (e) reliable? Explain.

E15.2 In the data file **USMacro_Monthly**, you will find data on two aggregate price series for the United States: the Consumer Price Index (CPI) and the Personal Consumption Expenditures Deflator (PCED). These series are alternative measures of consumer prices in the United States. The CPI prices a basket of goods whose composition is updated every 5–10 years. The PCED uses chain-weighting to price a basket of goods whose composition changes from month to month. Economists have argued that the CPI will overstate inflation because it does not take into account the substitution that occurs when relative prices change. If this substitution bias is important, then average CPI inflation should be systematically higher than PCED inflation. Let $\pi_t^{CPI} = 1200 \times \ln[CPI(t)/CPI(t-1)]$, $\pi_t^{PCED} = 1200 \times \ln[PCED(t)/PCED(t-1)]$, and $Y_t = \pi_t^{CPI} - \pi_t^{PCED}$, so π_t^{CPI} is the monthly rate of price inflation (measured in percentage points at an annual rate) based on the CPI, π_t^{PCED} is the monthly rate of price inflation from the PCED, and Y_t is the difference. Using data from 1970:1 through 2009:12, carry out the following exercises.

a. Compute the sample means of π_t^{CPI} and π_t^{PCED}. Are these point estimates consistent with the presence of economically significant substitution bias in the CPI?

b. Compute the sample mean of Y_t. Explain why it is numerically equal to the difference in the means computed in (a).

c. Show that the population mean of Y is equal to the difference of the population means of the two inflation rates.

d. Consider the "constant-term-only" regression: $Y_t = \beta_0 + u_t$. Show that $\beta_0 = E(Y)$. Do you think that u_t is serially correlated? Explain.

e. Construct a 95% confidence interval for β_0. What value of the HAC standard truncation parameter m did you choose? Why?

f. Is there statistically significant evidence that the mean inflation rate for the CPI is greater than the rate for the PCED?

g. Is there evidence of instability in β_0? Carry out a QLR test.

APPENDIX

15.1 The Orange Juice Data Set

The orange juice price data are the frozen orange juice component of processed foods and feeds group of the Producer Price Index (PPI), collected by the U.S. Bureau of Labor Statistics (BLS series wpu02420301). The orange juice price series was divided by the overall PPI for finished goods to adjust for general price inflation. The freezing degree days series was constructed from daily minimum temperatures recorded at Orlando-area airports, obtained from the National Oceanic and Atmospheric Administration (NOAA) of the U.S. Department of Commerce. The *FDD* series was constructed so that its timing and the timing of the orange juice price data were approximately aligned. Specifically, the frozen orange juice price data are collected by surveying a sample of producers in the middle of every month, although the exact date varies from month to month. Accordingly, the *FDD* series was constructed to be the number of freezing degree days from the 11th of one month to the 10th of the next month; that is, *FDD* is the maximum of zero and 32 minus the minimum daily temperature, summed over all days from the 11th to the 10th. Thus $\%ChgP_t$ for February is the percentage change in real orange juice prices from mid-January to mid-February, and FDD_t for February is the number of freezing degree days from January 11 to February 10.

APPENDIX

15.2 The ADL Model and Generalized Least Squares in Lag Operator Notation

This appendix presents the distributed lag model in lag operator notation, derives the ADL and quasi-differenced representations of the distributed lag model, and discusses the conditions under which the ADL model can have fewer parameters than the original distributed lag model.

The Distributed Lag, ADL, and Quasi-Differenced Models, in Lag Operator Notation

As defined in Appendix 14.3, the lag operator, L, has the property that $L^jX_t = X_{t-j}$, and the distributed lag $\beta_1X_t + \beta_2X_{t-1} + \cdots + \beta_{r+1}X_{t-r}$ can be expressed as $\beta(L)X_t$, where $\beta(L) = \sum_{j=0}^{r}\beta_{j+1}L^j$, where $L^0 = 1$. Thus the distributed lag model in Key Concept 15.1 [Equation (15.4)] can be written in lag operator notation as

$$Y_t = \beta_0 + \beta(L)X_t + u_t. \tag{15.40}$$

In addition, if the error term u_t follows an AR(p), then it can be written as

$$\phi(L)u_t = \tilde{u}_t, \tag{15.41}$$

where $\phi(L) = \sum_{j=0}^{p}\phi_jL^j$, where $\phi_0 = 1$ and $\tilde{u}_t$ is serially uncorrelated [note that $\phi_1, \ldots, \phi_p$ as defined here are the negatives of $\phi_1, \ldots, \phi_p$ in the notation of Equation (15.31)].

To derive the ADL model, premultiply each side of Equation (15.40) by $\phi(L)$ so that

$$\phi(L)Y_t = \phi(L)[\beta_0 + \beta(L)X_t + u_t] = \alpha_0 + \delta(L)X_t + \tilde{u}_t, \tag{15.42}$$

where

$$\alpha_0 = \phi(1)\beta_0 \text{ and } \delta(L) = \phi(L)\beta(L), \text{ where } \phi(1) = \sum_{j=0}^{p}\phi_j. \tag{15.43}$$

To derive the quasi-differenced model, note that $\phi(L)\beta(L)X_t = \beta(L)\phi(L)X_t = \beta(L)\tilde{X}_t$, where $\tilde{X}_t = \phi(L)X_t$. Thus rearranging Equation (15.42) yields

$$\tilde{Y}_t = \alpha_0 + \beta(L)\tilde{X}_t + \tilde{u}_t, \tag{15.44}$$

where $\tilde{Y}_t$ is the quasi-difference of Y_t; that is, $\tilde{Y}_t = \phi(L)Y_t$.

The ADL and GLS Estimators

The OLS estimator of the ADL coefficients is obtained by OLS estimation of Equation (15.42). The original distributed lag coefficients are $\beta(L)$, which, in terms of the estimated coefficients, is $\beta(L) = \delta(L)/\phi(L)$; that is, the coefficients in $\beta(L)$ satisfy the restrictions implied by $\phi(L)\beta(L) = \delta(L)$. Thus the estimator of the dynamic multipliers based on the OLS estimators of the coefficients of the ADL model, $\hat{\delta}(L)$ and $\hat{\phi}(L)$, is

$$\hat{\beta}^{ADL}(L) = \frac{\hat{\delta}(L)}{\hat{\phi}(L)}. \tag{15.45}$$

The expressions for the coefficients in Equation (15.29) in the text are obtained as a special case of Equation (15.45) when $r = 1$ and $p = 1$.

The feasible GLS estimator is computed by obtaining a preliminary estimator of $\phi(L)$, computing estimated quasi-differences, estimating $\beta(L)$ in Equation (15.44) using these estimated quasi-differences, and (if desired) iterating until convergence. The iterated GLS estimator is the NLLS estimator computed by NLLS estimation of the ADL model in Equation (15.42), subject to the nonlinear restrictions on the parameters contained in Equation (15.43).

As stressed in the discussion surrounding Equation (15.36) in the text, it is not enough for X_t to be (past and present) exogenous to use either of these estimation methods, for exogeneity alone does not ensure that Equation (15.36) holds. If, however, X is strictly exogenous, then Equation (15.36) does hold, and, assuming that Assumptions 2 through 4 of Key Concept 14.6 hold, these estimators are consistent and asymptotically normal. Moreover, the usual (cross-sectional heteroskedasticity-robust) OLS standard errors provide a valid basis for statistical inference.

Parameter reduction using the ADL model. Suppose that the distributed lag polynomial $\beta(L)$ can be written as a ratio of lag polynomials, $\theta_1(L)/\theta_2(L)$, where $\theta_1(L)$ and $\theta_2(L)$ are both lag polynomials of a low degree. Then $\phi(L)\beta(L)$ in Equation (15.43) is $\phi(L)\beta(L) = \phi(L)\theta_1(L)/\theta_2(L) = [\phi(L)/\theta_2(L)]\theta_1(L)$. If it so happens that $\phi(L) = \theta_2(L)$, then $\delta(L) = \phi(L)\beta(L) = \theta_1(L)$. If the degree of $\theta_1(L)$ is low, then q, the number of lags of X_t in the ADL model, can be much less than r. Thus, under these assumptions, estimation of the ADL model entails estimating potentially many fewer parameters than the original distributed lag model. It is in this sense that the ADL model can achieve more parsimonious parameterizations (that is, use fewer unknown parameters) than the distributed lag model.

As developed here, the assumption that $\phi(L)$ and $\theta_2(L)$ happen to be the same seems like a coincidence that would not occur in an application. However, the ADL model is able to capture a large number of shapes of dynamic multipliers with only a few coefficients.

ADL or GLS: Bias versus variance. A good way to think about whether to estimate dynamic multipliers by first estimating an ADL model and then computing the dynamic multipliers from the ADL coefficients or, alternatively, by estimating the distributed lag model directly using GLS is to view the decision in terms of a trade-off between bias and variance. Estimating the dynamic multipliers using an approximate ADL model introduces bias; however, because there are few coefficients, the variance of the estimator of the dynamic multipliers can be small. In contrast, estimating a long distributed lag model using GLS produces less bias in the multipliers; however, because there are so many coefficients, their variance can be large. If the ADL approximation to the dynamic multipliers is a good one, then the bias of the implied dynamic multipliers will be small, so the ADL approach will have a smaller variance than the GLS approach with only a small increase in the bias. For this reason, unrestricted estimation of an ADL model with small number of lags of Y and X is an attractive way to approximate a long distributed lag when X is strictly exogenous.

CHAPTER 16

Additional Topics in Time Series Regression

This chapter takes up some further topics in time series regression, starting with forecasting. Chapter 14 considered forecasting a single variable. In practice, however, you might want to forecast two or more variables such as the rate of inflation and the growth rate of the GDP. Section 16.1 introduces a model for forecasting multiple variables, vector autoregressions (VARs), in which lagged values of two or more variables are used to forecast future values of those variables. Chapter 14 also focused on making forecasts one period (e.g., one quarter) into the future, but making forecasts two, three, or more periods into the future is important as well. Methods for making multiperiod forecasts are discussed in Section 16.2.

Sections 16.3 and 16.4 return to the topic of Section 14.6, stochastic trends. Section 16.3 introduces additional models of stochastic trends and an alternative test for a unit autoregressive root. Section 16.4 introduces the concept of cointegration, which arises when two variables share a common stochastic trend, that is, when each variable contains a stochastic trend, but a weighted difference of the two variables does not.

In some time series data, especially financial data, the variance changes over time: Sometimes the series exhibits high volatility, while at other times the volatility is low, so the data exhibit clusters of volatility. Section 16.5 discusses volatility clustering and introduces models in which the variance of the forecast error changes over time, that is, models in which the forecast error is conditionally heteroskedastic. Models of conditional heteroskedasticity have several applications. One application is computing forecast intervals, where the width of the interval changes over time to reflect periods of high or low uncertainty. Another application is forecasting the uncertainty of returns on an asset, such as a stock, which in turn can be useful in assessing the risk of owning that asset.

16.1 Vector Autoregressions

Chapter 14 focused on forecasting the rate of inflation, but in reality economic forecasters are in the business of forecasting other key macroeconomic variables as well, such as the rate of unemployment, the growth rate of GDP, and interest