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

KEY CONCEPT 16.1

Vector Autoregressions

A vector autoregression (VAR) is a set of k time series regressions, in which the regressors are lagged values of all k series. A VAR extends the univariate autoregression to a list, or "vector," of time series variables. When the number of lags in each of the equations is the same and is equal to p, the system of equations is called a VAR(p).

In the case of two time series variables, Y_t and X_t, the VAR(p) consists of the two equations

$$Y_t = \beta_{10} + \beta_{11}Y_{t-1} + \cdots + \beta_{1p}Y_{t-p} + \gamma_{11}X_{t-1} + \cdots + \gamma_{1p}X_{t-p} + u_{1t} \quad (16.1)$$

$$X_t = \beta_{20} + \beta_{21}Y_{t-1} + \cdots + \beta_{2p}Y_{t-p} + \gamma_{21}X_{t-1} + \cdots + \gamma_{2p}X_{t-p} + u_{2t}, \quad (16.2)$$

where the β's and the γ's are unknown coefficients and u_{1t} and u_{2t} are error terms.

The VAR assumptions are the time series regression assumptions of Key Concept 14.6, applied to each equation. The coefficients of a VAR are estimated by estimating each equation by OLS.

rates. One approach is to develop a separate forecasting model for each variable using the methods of Section 14.4. Another approach is to develop a single model that can forecast all the variables, which can help to make the forecasts mutually consistent. One way to forecast several variables with a single model is to use a vector autoregression (VAR). A VAR extends the univariate autoregression to multiple time series variables, that is, it extends the univariate autoregression to a "vector" of time series variables.

The VAR Model

A **vector autoregression (VAR)** with two time series variables, Y_t and X_t, consists of two equations: In one, the dependent variable is Y_t; in the other, the dependent variable is X_t. The regressors in both equations are lagged values of both variables. More generally, a VAR with k time series variables consists of k equations, one for each of the variables; where the regressors in all equations are lagged values of all the variables. The coefficients of the VAR are estimated by estimating each of the equations by OLS.

VARs are summarized in Key Concept 16.1.

Inference in VARs. Under the VAR assumptions, the OLS estimators are consistent and have a joint normal distribution in large samples. Accordingly, statistical inference proceeds in the usual manner; for example, 95% confidence intervals on coefficients can be constructed as the estimated coefficient ± 1.96 standard errors.

One new aspect of hypothesis testing arises in VARs because a VAR with k variables is a collection, or system, of k equations. Thus it is possible to test joint hypotheses that involve restrictions across multiple equations.

For example, in the two-variable VAR(p) in Equations (16.1) and (16.2), you could ask whether the correct lag length is p or $p-1$; that is, you could ask whether the coefficients on Y_{t-p} and X_{t-p} are zero in these two equations. The null hypothesis that these coefficients are zero is

$$H_0: \beta_{1p} = 0, \beta_{2p} = 0, \gamma_{1p} = 0, \text{ and } \gamma_{2p} = 0. \quad (16.3)$$

The alternative hypothesis is that at least one of these four coefficients is nonzero. Thus the null hypothesis involves coefficients from *both* of the equations, two from each equation.

Because the estimated coefficients have a jointly normal distribution in large samples, it is possible to test restrictions on these coefficients by computing an F-statistic. The precise formula for this statistic is complicated because the notation must handle multiple equations, so we omit it. In practice, most modern software packages have automated procedures for testing hypotheses on coefficients in systems of multiple equations.

How many variables should be included in a VAR? The number of coefficients in each equation of a VAR is proportional to the number of variables in the VAR. For example, a VAR with five variables and four lags will have 21 coefficients (four lags each of five variables, plus the intercept) in each of the five equations, for a total of 105 coefficients! Estimating all these coefficients increases the amount of estimation error entering a forecast, which can result in a deterioration of the accuracy of the forecast.

The practical implication is that one needs to keep the number of variables in a VAR small and, especially, to make sure that the variables are plausibly related to each other so that they will be useful for forecasting one another. For example, we know from a combination of empirical evidence (such as that discussed in Chapter 14) and economic theory that the inflation rate, the unemployment rate, and the short-term interest rate are related to one another, suggesting that these variables could help to forecast one another other in a VAR. Including an unrelated variable in a VAR, however, introduces estimation error without adding predictive content, thereby reducing forecast accuracy.

Determining lag lengths in VARs.[1] Lag lengths in a VAR can be determined using either F-tests or information criteria.

The information criterion for a system of equations extends the single-equation information criterion in Section 14.5. To define this information criterion we need to adopt matrix notation. Let Σ_u be the $k \times k$ covariance matrix of the VAR errors and let $\hat{\Sigma}_u$ be the estimate of the covariance matrix where the i,j element of $\hat{\Sigma}_u$ is $\frac{1}{T}\sum_{t=1}^{T}\hat{u}_{it}\hat{u}_{jt}$, where $\hat{u}_{it}$ is the OLS residual from the i^{th} equation and $\hat{u}_{jt}$ is the OLS residual from the j^{th} equation. The BIC for the VAR is

$$\text{BIC}(p) = \ln[\det(\hat{\Sigma}_u)] + k(kp+1)\frac{\ln(T)}{T}, \tag{16.4}$$

where $\det(\hat{\Sigma}_u)$ is the determinant of the matrix $\hat{\Sigma}_u$. The AIC is computed using Equation (16.4), modified by replacing the term "ln(T)" by "2".

The expression for the BIC for the k equations in the VAR in Equation (16.4) extends the expression for a single equation given in Section 14.5. When there is a single equation, the first term simplifies to $\ln[SSR(p)/T]$. The second term in Equation (16.4) is the penalty for adding additional regressors; $k(kp+1)$ is the total number of regression coefficients in the VAR (there are k equations, each of which has an intercept and p lags of each of the k time series variables).

Lag length estimation in a VAR using the BIC proceeds analogously to the single equation case: Among a set of candidate values of p, the estimated lag length $\hat{p}$ is the value of p that minimizes BIC(p).

Using VARs for causal analysis. The discussion so far has focused on using VARs for forecasting. Another use of VAR models is for analyzing causal relationships among economic time series variables; indeed, it was for this purpose that VARs were first introduced to economics by the econometrician and macroeconomist Christopher Sims (1980). The use of VARs for causal inference is known as structural VAR modeling, "structural" because in this application VARs are used to model the underlying structure of the economy. Structural VAR analysis uses the techniques introduced in this section in the context of forecasting, plus some additional tools. The biggest conceptual difference between using VARs for forecasting and using them for structural modeling, however, is that structural

[1]This section uses matrices and may be skipped for less mathematical treatments.

modeling requires very specific assumptions, derived from economic theory and institutional knowledge, of what is exogenous and what is not. The discussion of structural VARs is best undertaken in the context of estimation of systems of simultaneous equations, which goes beyond the scope of this book. For an introduction to using VARs for forecasting and policy analysis, see Stock and Watson (2001). For additional mathematical detail on structural VAR modeling, see Hamilton (1994) or Watson (1994).

A VAR Model of the Rates of Inflation and Unemployment

As an illustration, consider a two-variable VAR for the inflation rate, Inf_t, and the rate of unemployment, $Unemp_t$. As in Chapter 14, we treat the rate of inflation as having a stochastic trend, so it is appropriate to transform it by computing its first difference, ΔInf_t.

The VAR for ΔInf_t and $Unemp_t$ consists of two equations: one in which ΔInf_t is the dependent variable and one in which $Unemp_t$ is the dependent variable. The regressors in both equations are lagged values of ΔInf_t and $Unemp_t$. Because of the apparent break in the Phillips curve in the early 1980s found in Section 14.7 using the QLR test, the VAR is estimated using data from 1982:I to 2004:IV.

The first equation of the VAR is the inflation equation:

$$\begin{aligned}\widehat{\Delta Inf_t} = {} & \underset{(0.55)}{1.47} - \underset{(0.12)}{0.64}\Delta Inf_{t-1} - \underset{(0.10)}{0.64}\Delta Inf_{t-2} - \underset{(0.11)}{0.13}\Delta Inf_{t-3} - \underset{(0.09)}{0.13}\Delta Inf_{t-4} \\ & - \underset{(0.58)}{3.49}Unemp_{t-1} + \underset{(0.94)}{2.80}Unemp_{t-2} + \underset{(1.07)}{2.44}Unemp_{t-3} - \underset{(0.55)}{2.03}Unemp_{t-4}.\end{aligned} \tag{16.5}$$

The adjusted R^2 is $\bar{R}^2 = 0.44$.

The second equation of the VAR is the unemployment equation, in which the regressors are the same as in the inflation equation but the dependent variable is the unemployment rate:

$$\begin{aligned}\widehat{Unemp_t} = {} & \underset{(0.12)}{0.22} + \underset{(0.017)}{0.005}\Delta Inf_{t-1} + \underset{(0.018)}{0.004}\Delta Inf_{t-2} - \underset{(0.018)}{0.007}\Delta Inf_{t-3} - \underset{(0.014)}{0.003}\Delta Inf_{t-4} \\ & + \underset{(0.11)}{1.52}Unemp_{t-1} - \underset{(0.18)}{0.29}Unemp_{t-2} - \underset{(0.21)}{0.43}Unemp_{t-3} + \underset{(0.11)}{0.16}Unemp_{t-4}.\end{aligned} \tag{16.6}$$

The adjusted R^2 is $\bar{R}^2 = 0.982$.

Equations (16.5) and (16.6), taken together, are a VAR(4) model of the change in the rate of inflation, ΔInf_t, and the unemployment rate, $Unemp_t$.

These VAR equations can be used to perform Granger causality tests. The F-statistic testing the null hypothesis that the coefficients on $Unemp_{t-1}$, $Unemp_{t-2}$, $Unemp_{t-3}$, and $Unemp_{t-4}$ are zero in the inflation equation [Equation (16.5)] is 11.04, which has a p-value less than 0.001. Thus the null hypothesis is rejected, so we can conclude that the unemployment rate is a useful predictor of changes in inflation, given lags in inflation (that is, the unemployment rate Granger-causes changes in inflation). The F-statistic testing the hypothesis that the coefficients on the four lags of ΔInf_t are zero in the unemployment equation [Equation (16.6)] is 0.16, which has a p-value of 0.96. Thus the change in the inflation rate does not Granger-cause the unemployment rate at the 10% significance level.

Forecasts of the rates of inflation and unemployment one period ahead are obtained exactly as discussed in Section 14.4. The forecast of the change of inflation from 2004:IV to 2005:I, based on Equation (16.5), is $\widehat{\Delta Inf}_{2005:I|2004:IV} = -0.1$ percentage point. A similar calculation using Equation (16.6) gives a forecast of the unemployment rate in 2005:I based on data through 2004:IV of $\widehat{Unemp}_{2005:I|2004:IV} = 5.4\%$, very close to its actual value, $Unemp_{2005:I} = 5.3\%$.

16.2 Multiperiod Forecasts

The discussion of forecasting so far has focused on making forecasts one period in advance. Often, however, forecasters are called upon to make forecasts further into the future. This section describes two methods for making multiperiod forecasts. The usual method is to construct "iterated" forecasts, in which a one-period-ahead model is iterated forward one period at a time, in a way that is made precise in this section. The second method is to make "direct" forecasts by using a regression in which the dependent variable is the multiperiod variable that one wants to forecast. For reasons discussed at the end of this section, in most applications the iterated method is recommended over the direct method.

Iterated Multiperiod Forecasts

The essential idea of an iterated forecast is that a forecasting model is used to make a forecast one period ahead, for period $T+1$ using data through period T. The model then is used to make a forecast for date $T+2$ given the data through date T, where the forecasted value for date $T+1$ is treated as data for the purpose of making the forecast for period $T+2$. Thus the one-period-ahead forecast (which is also referred to as a one-step-ahead forecast) is used as an intermediate step to make the two-period-ahead forecast. This process repeats, or iterates, until the forecast is made for the desired forecast horizon h.

The iterated AR forecast method: AR(1). An iterated AR(1) forecast uses an AR(1) for the one-period-ahead model. For example, consider the first order autoregression for ΔInf_t [Equation (14.7)]:

$$\widehat{\Delta Inf}_t = \underset{(0.13)}{0.02} - \underset{(0.10)}{0.24}\Delta Inf_{t-1}. \tag{16.7}$$

The first step in computing the two-quarter-ahead forecast of $\Delta Inf_{2005:II}$ based on Equation (16.7) using data through 2004:IV is to compute the one-quarter-ahead forecast of $\Delta Inf_{2005:I}$ based on data through 2004:IV: $\widehat{\Delta Inf}_{2005:I|2004:IV} = 0.02 - 0.24\Delta Inf_{2004:IV} = 0.02 - 0.24 \times 1.9 = -0.4$. The second step is to substitute this forecast into Equation (16.7) so that $\widehat{\Delta Inf}_{2005:II|2004:IV} = 0.02 - 0.24\widehat{\Delta Inf}_{2005:I|2004:IV} = 0.02 - 0.24 \times (-0.4) = 0.1$. Thus, based on information through the fourth quarter of 2004, this forecast states that the rate of inflation will increase by 0.1 percentage point between the first and second quarters of 2005.

The iterated AR forecast method: AR(p). The iterated AR(1) strategy is extended to an AR(p) by replacing Y_{T+1} with its forecast, $\hat{Y}_{T+1|T}$, and then treating that forecast as data for the AR(p) forecast of Y_{T+2}. For example, consider the iterated two-period-ahead forecast of inflation based on the AR(4) model from Section 14.3 [Equation (14.13)]:

$$\widehat{\Delta Inf}_t = \underset{(0.12)}{0.02} - \underset{(0.09)}{0.26}\Delta Inf_{t-1} - \underset{(0.08)}{0.32}\Delta Inf_{t-2} + \underset{(0.08)}{0.16}\Delta Inf_{t-3} - \underset{(0.09)}{0.03}\Delta Inf_{t-4}. \tag{16.8}$$

The forecast of $\Delta Inf_{2005:I}$ based on data through 2004:IV using this AR(4), computed in Section 14.3, is $\widehat{\Delta Inf}_{2005:I|2004:IV} = 0.4$. Thus the two-quarter-ahead iterated forecast based on the AR(4) is $\widehat{\Delta Inf}_{2005:II|2004:IV} = 0.02 - 0.26\widehat{\Delta Inf}_{2005:I|2004:IV} - 0.32\Delta Inf_{2004:IV} + 0.16\Delta Inf_{2004:III} - 0.03\Delta Inf_{2004:II} = 0.02 - 0.26 \times 0.4 - 0.32 \times 1.9 + 0.16 \times (-2.8) - 0.08 \times 0.6 = -1.1$. According to this iterated AR(4) forecast, based on data through the fourth quarter of 2004, the rate of inflation is predicted to fall by 1.1 percentage points between the first and second quarters of 2005.

Iterated multivariate forecasts using an iterated VAR. Iterated multivariate forecasts can be computed using a VAR in much the same way as iterated univariate forecasts are computed using an autoregression. The main new feature of an iterated multivariate forecast is that the two-step-ahead (period $T+2$) forecast of one variable depends on the forecasts of all variables in the VAR in period $T+1$. For example, to compute the forecast of the change of inflation from period $T+1$ to period $T+2$ using a VAR with the variables ΔInf_t and $Unemp_t$, one must forecast both ΔInf_{T+1} and $Unemp_{T+1}$ using data through period T as an intermediate step in forecasting ΔInf_{T+2}. More generally, to compute multiperiod iterated VAR forecasts h periods ahead, it is necessary to compute forecasts of all variables for all intervening periods between T and $T+h$.

As an example, we will compute the iterated VAR forecast of $\Delta Inf_{2005:II}$ based on data through 2004:IV using the VAR(4) for ΔInf_t and $Unemp_t$ in Section 16.1 [Equations (16.5) and (16.6)]. The first step is to compute the one-quarter-ahead forecasts $\widehat{\Delta Inf}_{2005:I|2004:IV}$ and $\widehat{Unemp}_{2005:I|2004:IV}$ from that VAR. The forecast $\widehat{\Delta Inf}_{2005:I|2004:IV}$ based on Equation (16.5) was computed in Section 14.3 and is -0.1 percentage point [Equation (14.18)]. A similar calculation using Equation (16.6) shows that $\widehat{Unemp}_{2005:I|2004:IV} = 5.4\%$. In the second step, these forecasts are substituted into Equations (16.5) and (16.6) to produce the two-quarter-ahead forecast, $\widehat{\Delta Inf}_{2005:II|2004:IV}$:

$$\begin{aligned}\widehat{\Delta Inf}_{2005:II|2004:IV} &= 1.47 - 0.64\widehat{\Delta Inf}_{2005:I|2004:IV} - 0.64\Delta Inf_{2004:IV} - 0.13\Delta Inf_{2004:III} \\ &\quad - 0.13\Delta Inf_{2004:II} - 3.49\widehat{Unemp}_{2005:I|2004:IV} + 2.80 Unemp_{2004:IV} \\ &\quad + 2.44 Unemp_{2004:III} - 2.03 Unemp_{2004:II} \\ &= 1.47 - 0.64 \times (-0.1) - 0.64 \times 1.9 - 0.13 \times (-2.8) - 0.13 \times 0.6 \\ &\quad - 3.49 \times 5.4 + 2.80 \times 5.4 + 2.44 \times 5.4 - 2.03 \times 5.6 = -1.1. \end{aligned} \tag{16.9}$$

Thus the iterated VAR(4) forecast, based on data through the fourth quarter of 2004, is that inflation will decline by 1.1 percentage points between the first and second quarters of 2005.

Iterated multiperiod forecasts are summarized in Key Concept 16.2.

Direct Multiperiod Forecasts

Direct multiperiod forecasts are computed without iterating by using a single regression in which the dependent variable is the multiperiod-ahead variable to be forecasted and the regressors are the predictor variables. Forecasts computed this way are called direct forecasts because the regression coefficients can be used directly to make the multiperiod forecast.

Iterated Multiperiod Forecasts

KEY CONCEPT 16.2

The **iterated multiperiod AR forecast** is computed in steps: First compute the one-period-ahead forecast, then use that to compute the two-period-ahead forecast, and so forth. The two- and three-period-ahead iterated forecasts based on an AR(p) are

$$\hat{Y}_{T+2|T} = \hat{\beta}_0 + \hat{\beta}_1 \hat{Y}_{T+1|T} + \hat{\beta}_2 Y_T + \hat{\beta}_3 Y_{T-1} + \cdots + \hat{\beta}_p Y_{T-p+2} \tag{16.10}$$

$$\hat{Y}_{T+3|T} = \hat{\beta}_0 + \hat{\beta}_1 \hat{Y}_{T+2|T} + \hat{\beta}_2 \hat{Y}_{T+1|T} + \hat{\beta}_3 Y_T + \cdots + \hat{\beta}_p Y_{T-p+3}, \tag{16.11}$$

where the $\hat{\beta}$'s are the OLS estimates of the AR(p) coefficients. Continuing this process ("iterating") produces forecasts further into the future.

The **iterated multiperiod VAR forecast** is also computed in steps: First compute the one-period-ahead forecast of all the variables in the VAR, then use those forecasts to compute the two-period-ahead forecasts, and continue this process iteratively to the desired forecast horizon. The two-period-ahead iterated forecast of Y_{T+2} based on the two-variable VAR(p) in Key Concept 16.1 is

$$\begin{aligned}\hat{Y}_{T+2|T} &= \hat{\beta}_{10} + \hat{\beta}_{11} \hat{Y}_{T+1|T} + \hat{\beta}_{12} Y_T + \hat{\beta}_{13} Y_{T-1} + \cdots + \hat{\beta}_{1p} Y_{T-p+2} \\ &\quad + \hat{\gamma}_{11} \hat{X}_{T+1|T} + \hat{\gamma}_{12} X_T + \hat{\gamma}_{13} X_{T-1} + \cdots + \hat{\gamma}_{1p} X_{T-p+2}, \end{aligned} \tag{16.12}$$

where the coefficients in Equation (16.12) are the OLS estimates of the VAR coefficients. Iterating produces forecasts further into the future.

The direct multiperiod forecasting method. Suppose that you want to make a forecast of Y_{T+2} using data through time T. The direct multivariate method takes the ADL model as its starting point, but lags the predictor variables by an additional time period. For example, if two lags of the predictors are used, then the dependent variable is Y_t and the regressors are Y_{t-2}, Y_{t-3}, X_{t-2}, and X_{t-3}. The coefficients from this regression can be used directly to compute the forecast of Y_{T+2} using data on Y_T, Y_{T-1}, X_T, and X_{T-1}, without the need for any iteration. More generally, in a direct h-period-ahead forecasting regression, all predictors are lagged h periods to produce the h-period-ahead forecast.

For example, the forecast of ΔInf_t two quarters ahead using four lags each of ΔInf_{t-2} and $Unemp_{t-2}$ is computed by first estimating the regression:

$$\widehat{\Delta Inf}_{t|t-2} = \underset{(0.53)}{-0.15} - \underset{(0.13)}{0.25}\Delta Inf_{t-2} + \underset{(0.13)}{0.16}\Delta Inf_{t-3} - \underset{(0.14)}{0.15}\Delta Inf_{t-4} - \underset{(0.07)}{0.10}\Delta Inf_{t-5}$$

$$- \underset{(0.70)}{0.17}Unemp_{t-2} + \underset{(1.63)}{1.82}Unemp_{t-3} - \underset{(2.00)}{3.53}Unemp_{t-4} + \underset{(0.91)}{1.89}Unemp_{t-5}. \quad (16.13)$$

The two-quarter-ahead forecast of the change of inflation from 2005:I to 2005:II is computed by substituting the values of $\Delta Inf_{2004:IV}, \ldots, \Delta Inf_{2004:I}, \ldots, Unemp_{2004:IV}, \ldots, Unemp_{2004:I}$ into Equation (16.13); this yields

$$\widehat{\Delta Inf}_{2005:II|2004:IV} = 0.15 - 0.25\Delta Inf_{2004:IV} + 0.16\Delta Inf_{2004:III} - 0.15\Delta Inf_{2004:II}$$
$$- 0.10\Delta Inf_{2004:I} - 0.17Unemp_{2004:IV} + 1.82Unemp_{2004:III}$$
$$- 3.53Unemp_{2004:II} + 1.89Unemp_{2004:I} = -1.38. \quad (16.14)$$

The three-quarter ahead direct forecast of ΔInf_{T+3} is computed by lagging all the regressors in Equation (16.13) by one additional quarter, estimating that regression, and then computing the forecast. The h-quarter-ahead direct forecast of ΔInf_{T+h} is computed by using ΔInf_t as the dependent variable and the regressors ΔInf_{t-h} and $Unemp_{t-h}$, plus additional lags of ΔInf_{t-h} and $Unemp_{t-h}$ as desired.

Standard errors in direct multiperiod regressions. Because the dependent variable in a multiperiod regression occurs two or more periods into the future, the error term in a multiperiod regression is serially correlated. To see this, consider the two-period-ahead forecast of inflation and suppose that a surprise jump in oil prices occurs in the next quarter. Today's two-period-ahead forecast of inflation will be too low because it does not incorporate this unexpected event. Because the oil price rise was also unknown in the previous quarter, the two-period-ahead forecast made last quarter will also be too low. Thus the surprise oil price jump next quarter means that *both* last quarter's and this quarter's two-period-ahead forecasts are too low. Because of such intervening events, the error term in a multiperiod regression is serially correlated.

As discussed in Section 15.4, if the error term is serially correlated, the usual OLS standard errors are incorrect or, more precisely, they are not a reliable basis for inference. Therefore, heteroskedasticity- and autocorrelation-consistent (HAC) standard errors must be used with direct multiperiod regressions. The standard errors reported in Equation (16.13) for direct multiperiod regressions therefore are Newey–West HAC standard errors, where the truncation parameter m is set according to Equation (15.17); for these data (for which $T = 92$), Equation (15.17) yields $m = 3$. For longer forecast horizons, the amount of overlap—and thus the degree of serial correlation in the error—increases: In general, the first $h - 1$ autocorrelation coefficients of the errors in an h-period-ahead regression are nonzero. Thus larger values of m than indicated by Equation (15.17) are appropriate for multiperiod regressions with long forecast horizons.

Direct multiperiod forcasts are summarized in Key Concept 16.3

Direct Multiperiod Forecasts

KEY CONCEPT 16.3

The **direct multiperiod forecast** h periods into the future based on p lags each of Y_t and an additional predictor X_t is computed by first estimating the regression,

$$Y_t = \delta_0 + \delta_1 Y_{t-h} + \cdots + \delta_p Y_{t-p-h+1} + \delta_{p+1} X_{t-h} + \cdots + \delta_{2p} X_{t-p-h+1} + u_t, \quad (16.15)$$

and then using the estimated coefficients directly to make the forecast of Y_{T+h} using data through period T.

Which Method Should You Use?

In most applications, the iterated method is the recommended procedure for multiperiod forecasting, for two reasons. First, from a theoretical perspective, if the underlying one-period-ahead model (the AR or VAR that is used to compute the iterated forecast) is specified correctly, then the coefficients are estimated more efficiently if they are estimated by a one-period-ahead regression (and then iterated) than by a multiperiod-ahead regression. Second, from a practical perspective, forecasters are usually interested in forecasts not just at a single horizon but at multiple horizons. Because they are produced using the same model, iterated forecasts tend to have time paths that are less erratic across horizons than do direct forecasts. Because a different model is used at every horizon for direct forecasts, sampling error in the estimated coefficients can add random fluctuations to the time paths of a sequence of direct multiperiod forecasts.

Under some circumstances, however, direct forecasts are preferable to iterated forecasts. One such circumstance is when you have reason to believe that the one-period-ahead model (the AR or VAR) is not specified correctly. For example, you might believe that the equation for the variable you are trying to forecast in a VAR is specified correctly, but that one or more of the other equations in the VAR is specified incorrectly, perhaps because of neglected nonlinear terms. If the one-step-ahead model is specified incorrectly, then in general the iterated multiperiod forecast will be biased and the MSFE of the iterated forecast can exceed the MSFE of the direct forecast, even though the direct forecast has a larger variance. A second circumstance in which a direct forecast might be desirable arises in multivariate forecasting models with many predictors, in which case a VAR specified in terms of all the variables could be unreliable because it would have very many estimated coefficients.

16.3 Orders of Integration and the DF-GLS Unit Root Test

This section extends the treatment of stochastic trends in Section 14.6 by addressing two further topics. First, the trends of some time series are not well described by the random walk model, so we introduce an extension of that model and discuss its implications for regression modeling of such series. Second, we continue the discussion of testing for a unit root in time series data and, among other things, introduce a second test for a unit root, the DF-GLS test.

Other Models of Trends and Orders of Integration

Recall that the random walk model for a trend, introduced in Section 14.6, specifies that the trend at date t equals the trend at date $t-1$, plus a random error term. If Y_t follows a random walk with drift β_0, then

$$Y_t = \beta_0 + Y_{t-1} + u_t, \tag{16.16}$$

where u_t is serially uncorrelated. Also recall from Section 14.6 that, if a series has a random walk trend, then it has an autoregressive root that equals 1.

Although the random walk model of a trend describes the long-run movements of many economic time series, some economic time series have trends that are smoother—that is, vary less from one period to the next—than is implied by Equation (16.16). A different model is needed to describe the trends of such series.

Orders of Integration, Differencing, and Stationarity

KEY CONCEPT 16.4

- If Y_t is integrated of order one, that is, if Y_t is $I(1)$, then Y_t has a unit autoregressive root and its first difference, ΔY_t, is stationary.
- If Y_t is integrated of order two, that is, if Y_t is $I(2)$, then ΔY_t has a unit autoregressive root and its second difference, $\Delta^2 Y_t$, is stationary.
- If Y_t is **integrated of order d**, that is, if Y_t is $I(d)$, then Y_t must be differenced d times to eliminate its stochastic trend, that is, $\Delta^d Y_t$ is stationary.

One model of a smooth trend makes the first difference of the trend follow a random walk; that is,

$$\Delta Y_t = \beta_0 + \Delta Y_{t-1} + u_t, \tag{16.17}$$

where u_t is serially uncorrelated. Thus, if Y_t follows Equation (16.17), ΔY_t follows a random walk, so $\Delta Y_t - \Delta Y_{t-1}$ is stationary. The difference of the first differences, $\Delta Y_t - \Delta Y_{t-1}$ is called the **second difference** of Y_t and is denoted $\Delta^2 Y_t = \Delta Y_t - \Delta Y_{t-1}$. In this terminology, if Y_t follows Equation (16.17), then its second difference is stationary. If a series has a trend of the form in Equation (16.17), then the first difference of the series has an autoregressive root that equals 1.

"Orders of integration" terminology. Some additional terminology is useful for distinguishing between these two models of trends. A series that has a random walk trend is said to be **integrated of order one**, or **$I(1)$**. A series that has a trend of the form in Equation (16.17) is said to be **integrated of order two**, or **$I(2)$**. A series that does not have a stochastic trend and is stationary is said to be **integrated of order zero**, or **$I(0)$**.

The **order of integration** in the $I(1)$ and $I(2)$ terminology is the number of times that the series needs to be differenced for it to be stationary: If Y_t is $I(1)$, then the first difference of Y_t, ΔY_t, is stationary, and if Y_t is $I(2)$, then the second difference of Y_t, $\Delta^2 Y_t$, is stationary. If Y_t is $I(0)$, then Y_t is stationary.

Orders of integration are summarized in Key Concept 16.4.

How to test whether a series is I(2) or I(1). If Y_t is $I(2)$, then ΔY_t is $I(1)$, so ΔY_t has an autoregressive root that equals 1. If, however, Y_t is $I(1)$, then ΔY_t is stationary. Thus the null hypothesis that Y_t is $I(2)$ can be tested against the alternative hypothesis that Y_t is $I(1)$ by testing whether ΔY_t has a unit autoregressive root. If the hypothesis that ΔY_t has a unit autoregressive root is rejected, then the hypothesis that Y_t is $I(2)$ is rejected in favor of the alternative that Y_t is $I(1)$.

Examples of I(2) and I(1) series: The price level and the rate of inflation. In Chapter 14, we concluded that the rate of inflation in the United States plausibly has a random walk stochastic trend, that is, that the rate of inflation is $I(1)$. If inflation is $I(1)$, then its stochastic trend is removed by first differencing, so ΔInf_t is stationary. Recall from Section 14.2 [Equation (14.2)] that quarterly inflation at an annual rate is the first difference of the logarithm of the price level, multiplied by 400; that is, $Inf_t = 400\Delta p_t$, where $p_t = \ln(CPI_t)$ and CPI_t denotes the value of the Consumer Price Index in quarter t. Thus treating the rate of inflation as $I(1)$ is equivalent to treating Δp_t as $I(1)$, but this in turn is equivalent to treating p_t as $I(2)$. Thus we have all along been treating the logarithm of the price level as $I(2)$, even though we have not used that terminology.

The logarithm of the price level, p_t, and the rate of inflation are plotted in Figure 16.1. The long-run trend of the logarithm of the price level (Figure 16.1a) varies more smoothly than the long-run trend in the rate of inflation (Figure 16.1b). The smoothly varying trend in the logarithm of the price level is typical of $I(2)$ series.

The DF-GLS Test for a Unit Root

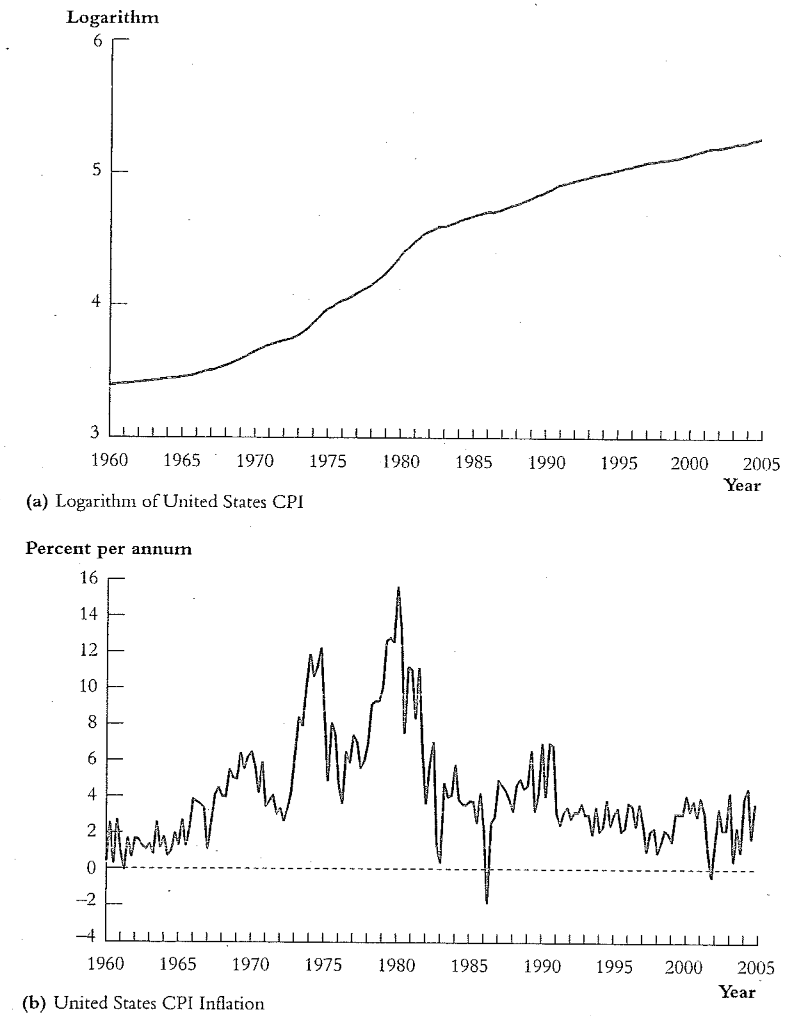
This section continues the discussion of Section 14.6 regarding testing for a unit autoregressive root. We first describe another test for a unit autoregressive root, the so-called DF-GLS test. Next, in an optional mathematical section, we discuss why unit root test statistics do not have normal distributions, even in large samples.

The DF-GLS test. The ADF test was the first test developed for testing the null hypothesis of a unit root and is the most commonly used test in practice. Other tests subsequently have been proposed, however, many of which have higher power (Key Concept 3.5) than the ADF test. A test with higher power than the ADF test is more likely to reject the null hypothesis of a unit root against the stationary alternative when the alternative is true; thus a more powerful test is better able to distinguish between a unit AR root and a root that is large but less than 1.

This section discusses one such test, the **DF-GLS test** developed by Elliott, Rothenberg, and Stock (1996). The test is introduced for the case that, under the null hypothesis, Y_t has a random walk trend, possibly with drift, and under the alternative Y_t is stationary around a linear time trend.

The DF-GLS test is computed in two steps. In the first step, the intercept and trend are estimated by generalized least squares (GLS; see Section 15.5). The GLS

FIGURE 16.1 The Logarithm of the Price Level and the Inflation Rate in the United States, 1960–2004



The trend in the logarithm of prices (Figure 16.1a) is much smoother than the trend in inflation (Figure 16.1b).

estimation is performed by computing three new variables, V_t, X_{1t}, and X_{2t}, where $V_1 = Y_1$ and $V_t = Y_t - \alpha^* Y_{t-1}$, $t = 2, \ldots, T$, $X_{11} = 1$ and $X_{1t} = 1 - \alpha^*$, $t = 2, \ldots, T$, and $X_{21} = 1$ and $X_{2t} = t - \alpha^*(t-1)$, where α^* is computed using the formula $\alpha^* = 1 - 13.5/T$. Then V_t is regressed against X_{1t} and X_{2t}; that is, OLS is used to estimate the coefficients of the population regression equation

$$V_t = \delta_0 X_{1t} + \delta_1 X_{2t} + e_t, \tag{16.18}$$

using the observations $t = 1, \ldots, T$, where e_t is the error term. Note that there is no intercept in the regression in Equation (16.18). The OLS estimators $\hat{\delta}_0$ and $\hat{\delta}_1$ are then used to compute a "detrended" version of Y_t, $Y_t^d = Y_t - (\hat{\delta}_0 + \hat{\delta}_1 t)$.

In the second step, the Dickey–Fuller test is used to test for a unit autoregressive root in Y_t^d, where the Dickey–Fuller regression does not include an intercept or a time trend. That is, ΔY_t^d is regressed against Y_{t-1}^d and $\Delta Y_{t-1}^d, \ldots, \Delta Y_{t-p}^d$, where the number of lags p is determined, as usual, either by expert knowledge or by using a data-based method such as the AIC or BIC as discussed in Section 14.5.

If the alternative hypothesis is that Y_t is stationary with a mean that might be nonzero but without a time trend, then the preceding steps are modified. Specifically, α^* is computed using the formula $\alpha^* = 1 - 7/T$, X_{2t} is omitted from the regression in Equation (16.18), and the series Y_t^d is computed as $Y_t^d = Y_t - \hat{\delta}_0$.

The GLS regression in the first step of the DF-GLS test makes this test more complicated than the conventional ADF test, but it is also what improves its ability to discriminate between the null hypothesis of a unit autoregressive root and the alternative that Y_t is stationary. This improvement can be substantial. For example, suppose that Y_t is in fact a stationary AR(1) with autoregressive coefficient $\beta_1 = 0.95$, that there are $T = 200$ observations, and that the unit root tests are computed without a time trend [that is, t is excluded from the Dickey–Fuller regression, and X_{2t} is omitted from Equation (16.18)]. Then the probability that the ADF test correctly rejects the null hypothesis at the 5% significance level is approximately 31% compared to 75% for the DF-GLS test.

Critical values for DF-GLS test. Because the coefficients on the deterministic terms are estimated differently in the ADF and DF-GLS tests, the tests have different critical values. The critical values for the DF-GLS test are given in Table 16.1. If the DF-GLS test statistic (the t-statistic on Y_{t-1}^d in the regression in the second step) is less than the critical value (that is, it is more negative than the critical value), then the null hypothesis that Y_t has a unit root is rejected. Like the critical values for the Dickey–Fuller test, the appropriate critical value depends on which version of the test is used, that is, on whether or not a time trend is included [whether or not X_{2t} is included in Equation (16.18)].

TABLE 16.1 Critical Values of the DF-GLS Test

Deterministic Regressors [Regressors in Equation (16.18)]	10%	5%	1%
Intercept only (X_{1t} only)	−1.62	−1.95	−2.58
Intercept and time trend (X_{1t} and X_{2t})	−2.57	−2.89	−3.48

Source: Fuller (1976) and Elliott, Rothenberg, and Stock (1996, Table 1).

Application to inflation. The DF-GLS statistic, computed for the rate of CPI inflation, Inf_t, over the period 1962:I to 2004:IV with an intercept but no time trend, is −2.06 when three lags of ΔY_t^d are included in the Dickey–Fuller regression in the second stage. This value is less than the 5% critical value in Table 16.1, −1.95, so using the DF-GLS test with three lags leads to rejecting the null hypothesis of a unit root at the 5% significance level. The choice of three lags was based on the AIC (out of a maximum of six lags).

Because the DF-GLS test is better able to discriminate between the unit root null hypothesis and the stationary alternative, one interpretation of this finding is that inflation is in fact stationary, but the Dickey–Fuller test implemented in Section 14.6 failed to detect this (at the 5% level). This conclusion, however, should be tempered by noting that whether the DF-GLS test rejects the null hypothesis is, in this application, sensitive to the choice of lag length. If the test is based on two lags, which is the number of lags chosen by BIC, it rejects the null hypothesis at the 10% but not the 5% level. The result is also sensitive to the choice of sample; if the statistic is instead computed over the period 1963:I to 2004:IV (that is, dropping just the first year), the test rejects the null hypothesis at the 10% level but not at the 5% level using AIC lag lengths. The overall picture therefore is rather ambiguous [as it is based on the ADF test, as discussed following Equation (14.34)] and requires the forecaster to make an informed judgment about whether it is better to model inflation as $I(1)$ or stationary.

Why Do Unit Root Tests Have Nonnormal Distributions?

In Section 14.6, it was stressed that the large-sample normal distribution upon which regression analysis relies so heavily does not apply if the regressors are nonstationary. Under the null hypothesis that the regression contains a unit root, the regressor Y_{t-1} in the Dickey–Fuller regression (and the regressor Y_{t-1}^d in the modified Dickey–Fuller regression in the second step of the DF-GLS test) is nonstationary. The nonnormal distribution of the unit root test statistics is a consequence of this nonstationarity.

To gain some mathematical insight into this nonnormality, consider the simplest possible Dickey–Fuller regression, in which ΔY_t is regressed against the single regressor Y_{t-1} and the intercept is excluded. In the notation of Key Concept 14.8, the OLS estimator in this regression is $\hat{\delta} = \sum_{t=1}^{T} Y_{t-1}\Delta Y_t / \sum_{t=1}^{T} Y_{t-1}^2$, so

$$T\hat{\delta} = \frac{\frac{1}{T}\sum_{t=1}^{T} Y_{t-1}\Delta Y_t}{\frac{1}{T^2}\sum_{t=1}^{T} Y_{t-1}^2}. \tag{16.19}$$

Consider the numerator in Equation (16.19). Under the additional assumption that $Y_0 = 0$, a bit of algebra (Exercise 16.5) shows that

$$\frac{1}{T}\sum_{t=1}^{T} Y_{t-1}\Delta Y_t = \frac{1}{2}\left[\left(\frac{Y_T}{\sqrt{T}}\right)^2 - \frac{1}{T}\sum_{t=1}^{T}(\Delta Y_t)^2\right]. \tag{16.20}$$

Under the null hypothesis, $\Delta Y_t = u_t$, which is serially uncorrelated and has a finite variance, so the second term in Equation (16.20) has the probability limit $\frac{1}{T}\sum_{t=1}^{T}(\Delta Y_t)^2 \xrightarrow{p} \sigma_u^2$. Under the assumption that $Y_0 = 0$, the first term in Equation (16.20) can be written $Y_T/\sqrt{T} = \sqrt{\frac{1}{T}}\sum_{t=1}^{T}\Delta Y_t = \sqrt{\frac{1}{T}}\sum_{t=1}^{T}u_t$, which in turn obeys the central limit theorem; that is, $Y_T/\sqrt{T} \xrightarrow{d} N(0, \sigma_u^2)$. Thus $(Y_T/\sqrt{T})^2 - \frac{1}{T}\sum_{t=1}^{T}(\Delta Y_t)^2 \xrightarrow{d} \sigma_u^2(Z^2 - 1)$, where Z is a standard normal random variable. Recall, however, that the square of a standard normal distribution has a chi-squared distribution with 1 degree of freedom. It therefore follows from Equation (16.20) that, under the null hypothesis, the numerator in Equation (16.19) has the limiting distribution

$$\frac{1}{T}\sum_{t=1}^{T} Y_{t-1}\Delta Y_t \xrightarrow{d} \frac{\sigma_u^2}{2}(\chi_1^2 - 1). \tag{16.21}$$

The large-sample distribution in Equation (16.21) is different than the usual large-sample normal distribution when the regressor is stationary. Instead, the numerator of the OLS estimator of the coefficient on Y_t in this Dickey–Fuller regression has a distribution that is proportional to a chi-squared distribution with 1 degree of freedom, minus 1.

This discussion has considered only the numerator of $T\hat{\delta}$. The denominator also behaves unusually under the null hypothesis: Because Y_t follows a random walk under the null hypothesis, $\frac{1}{T}\sum_{t=1}^{T} Y_{t-1}^2$ does not converge in probability to a constant. Instead, the denominator in Equation (16.19) is a random variable, even in large samples: Under the null hypothesis, $\frac{1}{T^2}\sum_{t=1}^{T} Y_{t-1}^2$ converges in distribution jointly with the numerator. The unusual distributions of the numerator and denominator in Equation (16.19) are the source of the nonstandard distribution of the Dickey–Fuller test statistic and the reason that the ADF statistic has its own special table of critical values.

16.4 Cointegration

Sometimes two or more series have the same stochastic trend in common. In this special case, referred to as cointegration, regression analysis can reveal long-run relationships among time series variables, but some new methods are needed.

Cointegration and Error Correction

Two or more time series with stochastic trends can move together so closely over the long run that they appear to have the same trend component; that is, they appear to have a **common trend**. For example, two interest rates on U.S. government debt are plotted in Figure 16.2. One of the rates is the interest rate on 90-day U.S. Treasury bills, at an annual rate ($R90_t$); the other is the interest rate on a 1-year U.S. Treasury bond ($R1yr_t$); these interest rates are discussed in Appendix 16.1. The interest rates exhibit the same long-run tendencies or trends: Both were low in the 1960s, both rose through the 1970s to peaks in the early 1980s, then both fell through the 1990s. Moreover, the difference between the two series, $R1yr_t - R90_t$, which is called the "spread" between the two interest rates and is also plotted in Figure 16.2, does not appear to have a trend. That is, subtracting the 90-day interest rate from the 1-year interest rate appears to eliminate the trends in both of the individual rates. Said differently, although the two interest rates differ, they appear to share a common stochastic trend: Because the trend in each individual series is eliminated by subtracting one series from the other, the two series must have the same trend; that is, they must have a common stochastic trend.

Two or more series that have a common stochastic trend are said to be cointegrated. The formal definition of **cointegration** (due to the econometrician Clive Granger, 1983; see the box on Clive Granger and Robert Engle) is given in Key Concept 16.5. In this section, we introduce a test for whether cointegration is present, discuss estimation of the coefficients of regressions relating cointegrated variables, and illustrate the use of the cointegrating relationship for forecasting.

KEY CONCEPT 16.5

Cointegration

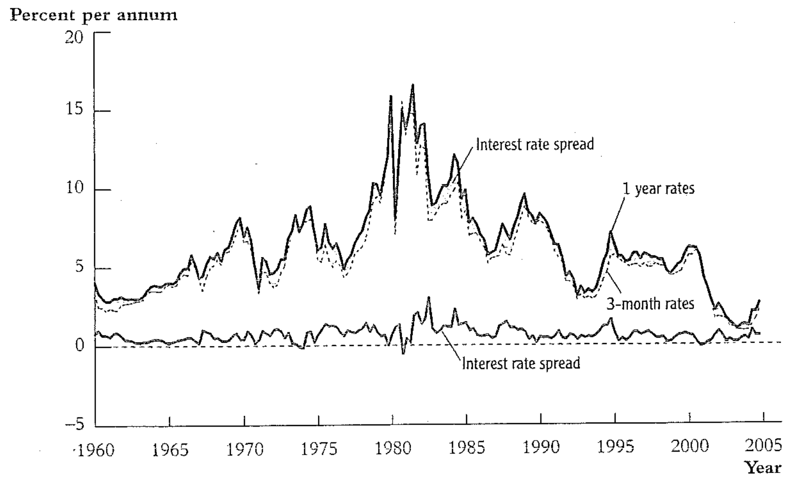
Suppose that X_t and Y_t are integrated of order one. If, for some coefficient θ, $Y_t - \theta X_t$ is integrated of order zero, then X_t and Y_t are said to be *cointegrated*. The coefficient θ is called the **cointegrating coefficient**.

If X_t and Y_t are cointegrated, then they have the same, or common, stochastic trend. Computing the difference $Y_t - \theta X_t$ eliminates this common stochastic trend.

The discussion initially focuses on the case that there are only two variables, X_t and Y_t.

Vector error correction model. Until now, we have eliminated the stochastic trend in an $I(1)$ variable Y_t by computing its first difference, ΔY_t; the problems created by stochastic trends were then avoided by using ΔY_t instead of Y_t in time series

FIGURE 16.2 One-Year Interest Rate, Three-Month Interest Rate, and Interest Rate Spread



One-year and three-month interest rates share a common stochastic trend. The spread, or the difference, between the two rates does not exhibit a trend. These two interest rates appear to be cointegrated.

regressions. If X_t and Y_t are cointegrated, however, another way to eliminate the trend is to compute $Y_t - \theta X_t$, where θ is chosen to eliminate the common trend from the difference. Because the term $Y_t - \theta X_t$ is stationary, it too can be used in regression analysis.

In fact, if X_t and Y_t are cointegrated, the first differences of X_t and Y_t can be modeled using a VAR, augmented by including $Y_{t-1} - \theta X_{t-1}$ as an additional regressor:

$$\Delta Y_t = \beta_{10} + \beta_{11}\Delta Y_{t-1} + \cdots + \beta_{1p}\Delta Y_{t-p} + \gamma_{11}\Delta X_{t-1} + \cdots + \gamma_{1p}\Delta X_{t-p} + \alpha_1(Y_{t-1} - \theta X_{t-1}) + u_{1t} \quad (16.22)$$

$$\Delta X_t = \beta_{20} + \beta_{21}\Delta Y_{t-1} + \cdots + \beta_{2p}\Delta Y_{t-p} + \gamma_{21}\Delta X_{t-1} + \cdots + \gamma_{2p}\Delta X_{t-p} + \alpha_2(Y_{t-1} - \theta X_{t-1}) + u_{2t}. \quad (16.23)$$

The term $Y_t - \theta X_t$ is called the **error correction term**. The combined model in Equations (16.22) and (16.23) is called a **vector error correction model** (VECM). In a VECM, past values of $Y_t - \theta X_t$ help to predict future values of ΔY_t and/or ΔX_t.

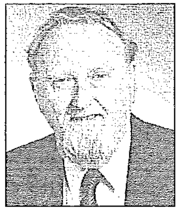
How Can You Tell Whether Two Variables Are Cointegrated?

There are three ways to decide whether two variables can plausibly be modeled as cointegrated: Use expert knowledge and economic theory, graph the series and see whether they appear to have a common stochastic trend, and perform statistical tests for cointegration. All three methods should be used in practice.

First, you must use your expert knowledge of these variables to decide whether cointegration is in fact plausible. For example, the two interest rates in Figure 16.2 are linked together by the so-called expectations theory of the term structure of interest rates. According to this theory, the interest rate on January 1 on the 1-year Treasury bond is the average of the interest rate on a 90-day Treasury bill for the first quarter of the year and the expected interest rates on future 90-day Treasury bills issued in the second, third, and fourth quarters of the year; if not, then investors could expect to make money by holding either the 1-year Treasury note or a sequence of four 90-day Treasury bills, and they would bid up prices until the expected returns are equalized. If the 90-day interest rate has a random walk stochastic trend, this theory implies that this stochastic trend is inherited by the 1-year interest rate and that the difference between the two rates, that is, the spread, is stationary. Thus the expectations theory of the term structure

Robert Engle and Clive Granger, Nobel Laureates

In 2003, two econometricians, Robert F. Engle and Clive W. J. Granger, won the Nobel Prize in economics for fundamental theoretical research in time series econometrics that they did in the late 1970s and early 1980s.

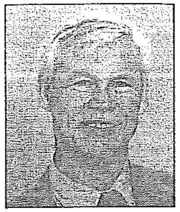


Clive W. J. Granger

Granger's work focused on how to handle stochastic trends in economic time series data. From earlier work by himself and others, he knew that two unrelated series with stochastic trends could, by the usual statistical measures of t-statistics and regression R^2's, falsely appear to be meaningfully related; this is the "spurious regression" problem. In the 1970s, the standard practice was to use differences of time series data to avoid the risk of a spurious regression. For this reason, Granger was skeptical of some recent work by some British econometricians (Davidson, Hendry, Srba, and Yeo, 1978), who claimed that the lagged difference between log consumption and log income ($\ln C_{t-1} - \ln Y_{t-1}$) was a valuable predictor of the growth rate of consumption ($\Delta \ln C_t$). Because $\ln C_t$ and $\ln Y_t$ individually have a unit root, the conventional wisdom was that they should be included in first differences because including them in levels would produce a version of a spurious regression.

Granger set out to prove mathematically that the British team had made a mistake, but instead proved that their specification was correct: There is a well-defined mathematical representation—the vector error correction model—for time series that are individually $I(1)$ but for which a linear combination is $I(0)$. He termed this situation "cointegration." In subsequent work with his colleague at the University of California at San Diego, Robert Engle, Granger proposed several tests for cointegration, most notably the Engle–Granger ADF test described on page 653. The methods of cointegration analysis are now a staple in modern macroeconometrics.

Around the same time, Robert Engle was pondering the striking increase in the volatility of U.S. inflation during the late 1970s (see Figure 16.1b). If the volatility of inflation had increased, he reasoned, then prediction intervals for inflation forecasts should be wider than the models of the day would indicate, because those models held the variance of inflation constant. But how, precisely, can you forecast the time-varying variance (which you do not observe) of an error term (which you also do not observe)?



Robert F. Engle

Engle's answer was to develop the autoregressive conditional heteroskedasticity (ARCH) model, described in Section 16.5. The ARCH model and its extensions, developed mainly by Engle and his students, proved especially useful for modeling the volatility of asset returns, and the resulting volatility forecasts can be used to price financial derivatives and to assess changes over time in the risk of holding financial assets. Today, measures and forecasts of volatility are a core component of financial econometrics, and the ARCH model and its descendants are the workhorse tools for modeling volatility.

implies that if the interest rates are $I(1)$, then they will be cointegrated with a cointegrating coefficient of $\theta = 1$ (Exercise 16.2).

Second, visual inspection of the series helps to identify cases in which cointegration is plausible. For example, the graph of the two interest rates in Figure 16.2 shows that each of the series appears to be $I(1)$ but that the spread appears to be $I(0)$, so the two series appear to be cointegrated.

Third, the unit root testing procedures introduced so far can be extended to tests for cointegration. The insight on which these tests are based is that if Y_t and X_t are cointegrated with cointegrating coefficient θ, then $Y_t - \theta X_t$ is stationary; otherwise, $Y_t - \theta X_t$ is nonstationary [is $I(1)$]. The hypothesis that Y_t and X_t are not cointegrated [that is, that $Y_t - \theta X_t$ is $I(1)$] therefore can be tested by testing the null hypothesis that $Y_t - \theta X_t$ has a unit root; if this hypothesis is rejected, then Y_t and X_t can be modeled as cointegrated. The details of this test depend on whether the cointegrating coefficient θ is known.

Testing for cointegration when θ is known. In some cases expert knowledge or economic theory suggests values of θ. When θ is known, the Dickey–Fuller and DF-GLS unit root tests can be used to test for cointegration by first constructing the series $z_t = Y_t - \theta X_t$ and then testing the null hypothesis that z_t has a unit autoregressive root.

Testing for cointegration when θ is unknown. If the cointegrating coefficient θ is unknown, then it must be estimated prior to testing for a unit root in the error correction term. This preliminary step makes it necessary to use different critical values for the subsequent unit root test.

Specifically, in the first step the cointegrating coefficient θ is estimated by OLS estimation of the regression

$$Y_t = \alpha + \theta X_t + z_t. \tag{16.24}$$

In the second step, a Dickey–Fuller t-test (with an intercept but no time trend) is used to test for a unit root in the residual from this regression, $\hat{z}_t$. This two-step procedure is called the Engle–Granger Augmented Dickey–Fuller test for cointegration, or **EG-ADF test** (Engle and Granger, 1987).

Critical values of the EG-ADF statistic are given in Table 16.2.[2] The critical values in the first row apply when there is a single regressor in Equation (16.26),

[2]The critical values in Table 16.2 are taken from Fuller (1976) and Phillips and Ouliaris (1990). Following a suggestion by Hansen (1992), the critical values in Table 16.2 are chosen so that they apply whether or not X_t and Y_t have drift components.

TABLE 16.2 Critical Values for the Engle–Granger ADF Statistic

Number of X's in Equation (16.24)	10%	5%	1%
1	−3.12	−3.41	−3.96
2	−3.52	−3.80	−4.36
3	−3.84	−4.16	−4.73
4	−4.20	−4.49	−5.07

so there are two cointegrated variables (X_t and Y_t). The subsequent rows apply to the case of multiple cointegrated variables, which is discussed at the end of this section.

Estimation of Cointegrating Coefficients

If X_t and Y_t are cointegrated, then the OLS estimator of the coefficient in the cointegrating regression in Equation (16.24) is consistent. However, in general the OLS estimator has a nonnormal distribution, and inferences based on its t-statistics can be misleading whether or not those t-statistics are computed using HAC standard errors. Because of these drawbacks of the OLS estimator of θ, econometricians have developed a number of other estimators of the cointegrating coefficient.

One such estimator of θ that is simple to use in practice is the **dynamic OLS (DOLS) estimator** (Stock and Watson, 1993). The DOLS estimator is based on a modified version of Equation (16.24) that includes past, present, and future values of the change in X_t:

$$Y_t = \beta_0 + \theta X_t + \sum_{j=-p}^{p} \delta_j \Delta X_{t-j} + u_t. \qquad (16.25)$$

Thus, in Equation (16.25), the regressors are $X_t, \Delta X_{t+p}, \ldots, \Delta X_{t-p}$. The DOLS estimator of θ is the OLS estimator of θ in the regression of Equation (16.25).

If X_t and Y_t are cointegrated, then the DOLS estimator is efficient in large samples. Moreover, statistical inferences about θ and the δ's in Equation (16.25) based on HAC standard errors are valid. For example, the t-statistic constructed using the DOLS estimator with HAC standard errors has a standard normal distribution in large samples.

One way to interpret Equation (16.25) is to recall from Section 15.3 that cumulative dynamic multipliers can be computed by modifying the distributed lag regression of Y_t on X_t and its lags. Specifically, in Equation (15.7), the cumulative dynamic multipliers were computed by regressing Y_t on ΔX_t, lags of ΔX_t, and X_{t-r}; the coefficient on X_{t-r} in that specification is the long-run cumulative dynamic multiplier. Similarly, if X_t were strictly exogenous, then in Equation (16.25) the coefficient on X_t, θ would be the long-run cumulative multiplier, that is, the long-run effect on Y of a change in X. If X_t is not strictly exogenous, then the coefficients do not have this interpretation. Nevertheless, because X_t and Y_t have a common stochastic trend if they are cointegrated, the DOLS estimator is consistent even if X_t is endogenous.

The DOLS estimator is not the only efficient estimator of the cointegrating coefficient. The first such estimator was developed by Søren Johansen (Johansen, 1988). For a discussion of Johansen's method and of other ways to estimate the cointegrating coefficient, see Hamilton (1994, Chapter 20).

Even if economic theory does not suggest a specific value of the cointegrating coefficient, it is important to check whether the estimated cointegrating relationship makes sense in practice. Because cointegration tests can be misleading (they can improperly reject the null hypothesis of no cointegration more frequently than they should, and frequently they improperly fail to reject the null), it is especially important to rely on economic theory, institutional knowledge, and common sense when estimating and using cointegrating relationships.

Extension to Multiple Cointegrated Variables

The concepts, tests, and estimators discussed here extend to more than two variables. For example, if there are three variables, Y_t, X_{1t}, and X_{2t}, each of which is $I(1)$, then they are cointegrated with cointegrating coefficients θ_1 and θ_2 if $Y_t - \theta_1 X_{1t} - \theta_2 X_{2t}$ is stationary. When there are three or more variables, there can be multiple cointegrating relationships. For example, consider modeling the relationship among three interest rates: the 3-month rate, the 1-year rate, and the 5-year rate ($R5yr$). If they are $I(1)$, then the expectations theory of the term structure of interest rates suggests that they will all be cointegrated. One cointegrating relationship suggested by the theory is $R1yr_t - R90_t$, and a second relationship is $R5yr_t - R90_t$. (The relationship $R5yr_t - R1yr_t$ is also a cointegrating relationship, but it contains no additional information beyond that in the other relationships because it is perfectly multicollinear with the other two cointegrating relationships.)

The EG-ADF procedure for testing for a single cointegrating relationship among multiple variables is the same as for the case of two variables, except that the regression in Equation (16.24) is modified so that both X_{1t} and X_{2t} are regressors; the critical values for the EG-ADF test are given in Table 16.2, where the appropriate row depends on the number of regressors in the first-stage OLS cointegrating regression. The DOLS estimator of a single cointegrating relationship among multiple X's involves including the level of each X along with leads and lags of the first difference of each X. Tests for multiple cointegrating relationships can be performed using system methods such as Johansen's (1988) method, and the DOLS estimator can be extended to multiple cointegrating relationships by estimating multiple equations, one for each cointegrating relationship. For additional discussion of cointegration methods for multiple variables, see Hamilton (1994).

A cautionary note. If two or more variables are cointegrated, then the error correction term can help to forecast these variables and, possibly, other related variables. However, cointegration requires the variables to have the same stochastic trends. Trends in economic variables typically arise from complex interactions of disparate forces, and closely related series can have different trends for subtle reasons. If variables that are not cointegrated are incorrectly modeled using a VECM, then the error correction term will be $I(1)$; this introduces a trend into the forecast that can result in poor out-of-sample forecast performance. Thus forecasting using a VECM must be based on a combination of compelling theoretical arguments in favor of cointegration and careful empirical analysis.

Application to Interest Rates

As discussed earlier, the expectations theory of the term structure of interest rates implies that if two interest rates of different maturities are $I(1)$, then they will be cointegrated with a cointegrating coefficient of $\theta = 1$; that is, the spread between the two rates will be stationary. Inspection of Figure 16.2 provides qualitative support for the hypothesis that the 1-year and 3-month interest rates are cointegrated. We first use unit root and cointegration test statistics to provide more formal evidence on this hypothesis, then estimate a vector error correction model for these two interest rates.

Unit root and cointegration tests. Various unit root and cointegration test statistics for these two series are reported in Table 16.3. The unit root test statistics in the first two rows examine the hypothesis that the two interest rates, the 3-month rate ($R90$) and the 1-year rate ($R1yr$), individually have a unit root. Two of the four statistics in the first two rows fail to reject this hypothesis at the 10% level, and three of the four fail to reject at the 5% level. The exception is the ADF statistic evaluated for the 90-day Treasury bill rate (–2.96), which rejects the unit root hypothesis at the 5% level. The ADF and DF-GLS statistics lead to different

TABLE 16.3 Unit Root and Cointegration Test Statistics for Two Interest Rates

Series	ADF Statistic	DF-GLS Statistic
$R90$	–2.96*	–1.88
$R1yr$	–2.22	–1.37
$R1yr - R90$	–6.31**	–5.59**
$R1yr - 1.046R90$	–6.97**	–

$R90$ is the interest rate on 90-day U.S. Treasury bills, at an annual rate, and $R1yr$ is the interest rate on 1-year U.S. Treasury bonds. Regressions were estimated using quarterly data over the period 1962:I–1999:IV. The number of lags in the unit root test statistic regressions were chosen by AIC (six lags maximum). Unit root test statistics are significant at the *5% or **1% significance level.

conclusions for this variable (the ADF test rejects the unit root hypothesis at the 5% level while the DF-GLS test does not), which means that we must exercise some judgment in deciding whether these variables are plausibly modeled as $I(1)$. Taken together, these results suggest that the interest rates are plausibly modeled as $I(1)$.

The unit root statistics for the spread, $R1yr_t - R90_t$, test the further hypothesis that these variables are not cointegrated against the alternative that they are. The null hypothesis that the spread contains a unit root is rejected at the 1% level using both unit root tests. Thus we reject the hypothesis that the series are not cointegrated against the alternative that they are, with a cointegrating coefficient $\theta = 1$. Taken together, the evidence in the first three rows of Table 16.3 suggests that these variables plausibly can be modeled as cointegrated with $\theta = 1$.

Because in this application economic theory suggests a value for θ (the expectations theory of the term structure suggests that $\theta = 1$) and because the error correction term is $I(0)$ when this value is imposed (the spread is stationary), in principle it is not necessary to use the EG-ADF test, in which θ is estimated. Nevertheless, we compute the test as an illustration. The first step in the EG-ADF test is to estimate θ by the OLS regression of one variable on the other; the result is

$$\widehat{R1yr_t} = 0.361 + 1.046R90_t, \quad \overline{R}^2 = 0.973. \tag{16.26}$$

The second step is to compute the ADF statistic for the residual from this regression, $\hat{z}_t$. The result, given in the final row of Table 16.3, is less than the 1% critical value of –3.96 in Table 16.2, so the null hypothesis that $\hat{z}_t$ has a unit autoregressive root is rejected. This statistic also points toward treating the two interest rates as cointegrated. Note that no standard errors are presented in Equation (16.26) because, as previously discussed, the OLS estimator of the cointegrating coefficient has a nonnormal distribution and its t-statistic is not normally distributed, so presenting standard errors (HAC or otherwise) would be misleading.

A vector error correction model of the two interest rates. If Y_t and X_t are cointegrated, then forecasts of ΔY_t and ΔX_t can be improved by augmenting a VAR of ΔY_t and ΔX_t by the lagged value of the error correction term, that is, by computing forecasts using the VECM in Equations (16.22) and (16.23). If θ is known, then the unknown coefficients of the VECM can be estimated by OLS, including $z_{t-1} = Y_{t-1} - \theta X_{t-1}$ as an additional regressor. If θ is unknown, then the VECM can be estimated using $\hat{z}_{t-1}$ as a regressor, where $\hat{z}_t = Y_t - \hat{\theta} X_t$, where $\hat{\theta}$ is an estimator of θ.

In the application to the two interest rates, theory suggests that $\theta = 1$, and the unit root tests support modeling the two interest rates as cointegrated with a cointegrating coefficient of 1. We therefore specify the VECM using the theoretically suggested value of $\theta = 1$, that is, by adding the lagged value of the spread, $R1yr_{t-1} - R90_{t-1}$, to a VAR in $\Delta R1yr_t$ and $\Delta R90_t$. Specified with two lags of first differences, the resulting VECM is

$$\begin{aligned}\widehat{\Delta R90_t} = \underset{(0.17)}{0.14} - \underset{(0.32)}{0.24}\Delta R90_{t-1} - \underset{(0.34)}{0.44}\Delta R90_{t-2} - \underset{(0.39)}{0.01}\Delta R1yr_{t-1}\\ + \underset{(0.27)}{0.15}\Delta R1yr_{t-2} - \underset{(0.27)}{0.18}(R1yr_{t-1} - R90_{t-1})\end{aligned} \quad (16.27)$$

$$\begin{aligned}\widehat{\Delta R1yr_t} = \underset{(0.16)}{0.36} - \underset{(0.30)}{0.14}\Delta R90_{t-1} - \underset{(0.29)}{0.33}\Delta R90_{t-2} - \underset{(0.35)}{0.11}\Delta R1yr_{t-1}\\ + \underset{(0.25)}{0.10}\Delta R1yr_{t-2} - \underset{(0.24)}{0.52}(R1yr_{t-1} - R90_{t-1})\end{aligned} \quad (16.28)$$

In the first equation, none of the coefficients is individually significant at the 5% level and the coefficients on the lagged first differences of the interest rates are not jointly significant at the 5% level. In the second equation, the coefficients on the lagged first differences are not jointly significant, but the coefficient on the lagged spread (the error correction term), which is estimated to be −0.52, has a t-statistic of −2.17, so it is statistically significant at the 5% level. Although lagged values of the first difference of the interest rates are not useful for predicting future

interest rates, the lagged spread does help to predict the change in the 1-year Treasury bond rate. When the 1-year rate exceeds the 90-day rate, the 1-year rate is forecasted to fall in the future.

16.5 Volatility Clustering and Autoregressive Conditional Heteroskedasticity

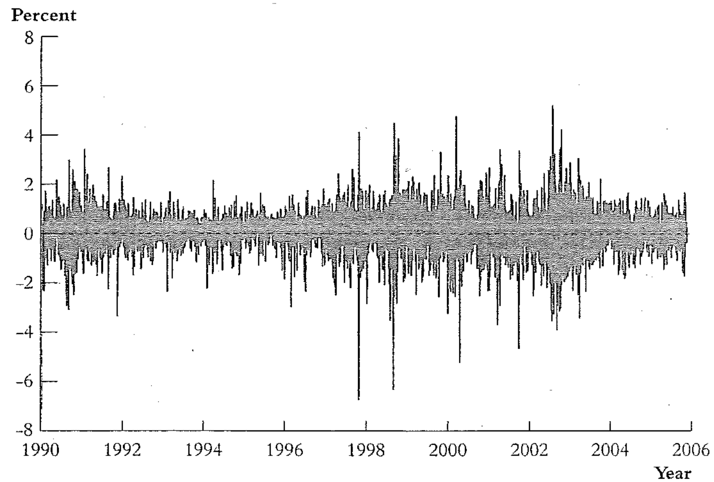
The phenomenon that some times are tranquil while others are not—that is, that volatility comes in clusters—shows up in many economic time series. This section presents a pair of models for quantifying volatility clustering or, as it is also known, conditional heteroskedasticity.

Volatility Clustering

The volatility of many financial and macroeconomic variables changes over time. For example, daily percentage changes in the New York Stock Exchange (NYSE) stock price index, shown in Figure 16.3, exhibit periods of high volatility, such as

FIGURE 16.3 Daily Percentage Changes in the NYSE Index, 1990–2005

Daily NYSE percentage price changes exhibit volatility clustering, in which there are some periods of high volatility, such as in the late 1990s, and other periods of relative tranquility, such as in the mid-1990s.



in 1990 and 2003, and other periods of low volatility, such as in 1993. A series with some periods of low volatility and some periods of high volatility is said to exhibit **volatility clustering**. Because the volatility appears in clusters, the variance of the daily percentage price change in the NYSE index can be forecasted, even though the daily price change itself is very difficult to forecast.

Forecasting the variance of a series is of interest for several reasons. First, the variance of an asset price is a measure of the risk of owning that asset: The larger the variance of daily stock price changes, the more a stock market participant stands to gain—or to lose—on a typical day. An investor who is worried about risk would be less tolerant of participating in the stock market during a period of high—rather than low—volatility.

Second, the value of some financial derivatives, such as options, depends on the variance of the underlying asset. An options trader wants the best available forecasts of future volatility to help him or her know the price at which to buy or sell options.

Third, forecasting variances makes it possible to have accurate forecast intervals. Suppose that you are forecasting the rate of inflation. If the variance of the forecast error is constant, then an approximate forecast confidence interval can be constructed along the lines discussed in Section 14.4—that is, as the forecast plus or minus a multiple of the *SER*. If, however, the variance of the forecast error changes over time, then the width of the forecast interval should change over time: At periods when inflation is subject to particularly large disturbances or shocks, the interval should be wide; during periods of relative tranquility, the interval should be tighter.

Volatility clustering can be thought of as clustering of the variance of the error term over time: If the regression error has a small variance in one period, its variance tends to be small in the next period, too. In other words, volatility clustering implies that the error exhibits time-varying heteroskedasticity.

Autoregressive Conditional Heteroskedasticity

Two models of volatility clustering are the **autoregressive conditional heteroskedasticity (ARCH)** model and its extension, the **generalized ARCH (GARCH)** model.

ARCH. Consider the ADL(1,1) regression

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \gamma_1 X_{t-1} + u_t. \tag{16.29}$$

In the ARCH model, which was developed by the econometrician Robert Engle (Engle, 1982; see the box on Clive Granger and Robert Engle), the error u_t is modeled as being normally distributed with mean zero and variance σ_t^2, where σ_t^2 depends on past squared values u_t. Specifically, the ARCH model of order p, denoted ARCH(p), is

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \alpha_2 u_{t-2}^2 + \cdots + \alpha_p u_{t-p}^2, \tag{16.30}$$

where $\alpha_0, \alpha_1, \ldots, \alpha_p$ are unknown coefficients. If these coefficients are positive, then if recent squared errors are large the ARCH model predicts that the current squared error will be large in magnitude in the sense that its variance, σ_t^2, is large.

Although it is described here for the ADL(1,1) model in Equation (16.29), the ARCH model can be applied to the error variance of any time series regression model with an error that has a conditional mean of zero, including higher-order ADL models, autoregressions, and time series regressions with multiple predictors.

GARCH. The generalized ARCH (GARCH) model, developed by the econometrician Tim Bollerslev (1986), extends the ARCH model to let σ_t^2 depend on its own lags as well as lags of the squared error. The GARCH(p,q) model is

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \cdots + \alpha_p u_{t-p}^2 + \phi_1 \sigma_{t-1}^2 + \cdots + \phi_q \sigma_{t-q}^2, \tag{16.31}$$

where $\alpha_0, \alpha_1, \ldots, \alpha_p, \phi_1, \ldots, \phi_q$ are unknown coefficients.

The ARCH model is analogous to a distributed lag model, and the GARCH model is analogous to an ADL model. As discussed in Appendix 15.2, the ADL model (when appropriate) can provide a more parsimonious model of dynamic multipliers than the distributed lag model. Similarly, by incorporating lags of σ_t^2, the GARCH model can capture slowly changing variances with fewer parameters than the ARCH model.

An important application of ARCH and GARCH models is to measuring and forecasting the time-varying volatility of returns on financial assets, particularly assets observed at high sampling frequencies such as the daily stock returns in Figure 16.3. In such applications the return itself is often modeled as unpredictable, so the regression in Equation (16.29) only includes the intercept.

Estimation and inference. ARCH and GARCH models are estimated by the method of maximum likelihood (Appendix 11.2). The estimators of the ARCH and GARCH coefficients are normally distributed in large samples, so in large samples t-statistics have standard normal distributions and confidence intervals can be constructed as the maximum likelihood estimate ± 1.96 standard errors.

Application to Stock Price Volatility

A GARCH(1,1) model of the NYSE daily percentage stock price changes, R_t, estimated using data on all trading days from January 2, 1990, through November 11, 2005, is

$$\hat{R}_t = \underset{(0.012)}{0.049} \qquad (16.32)$$

$$\hat{\sigma}_t^2 = \underset{(0.0014)}{0.0079} + \underset{(0.005)}{0.072}u_{t-1}^2 + \underset{(0.006)}{0.919}\sigma_{t-1}^2 \qquad (16.33)$$

No lagged predictors appear in Equation (16.32) because daily NYSE price changes are essentially unpredictable.

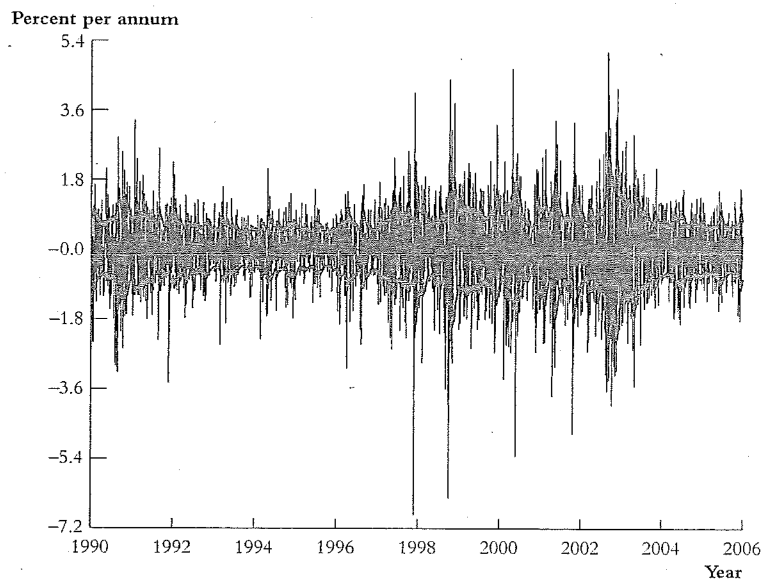
The two coefficients in the GARCH model (the coefficients on u_{t-1}^2 and σ_{t-1}^2) are both individually statistically significant at the 5% significance level. One measure of the persistence of movements in the variance is the sum of the coefficients on u_{t-1}^2 and σ_{t-1}^2 in the GARCH model (Exercise 16.9). This sum (0.991) is large, indicating that changes in the conditional variance are persistent. Said differently, the estimated GARCH model implies that periods of high volatility in NYSE prices will be long-lasting. This implication is consistent with the long periods of volatility clustering seen in Figure 16.3.

The estimated conditional variance at date t, $\hat{\sigma}_t^2$, can be computed using the residuals from Equation (16.32) and the coefficients in Equation (16.33). Figure 16.4 plots bands of plus or minus one conditional standard deviation (that is, $\pm\hat{\sigma}_t$) based on the GARCH(1,1) model, along with deviations of the percentage price change series from its mean. The conditional standard deviation bands quantify the time-varying volatility of the daily price changes. During the mid-1990s, the conditional standard deviation bands are tight, indicating lower levels of risk for investors holding the NYSE index. In contrast, around the turn of the century, these conditional standard deviation bands are wide, indicating a period of greater daily stock price volatility.

16.6 Conclusion

This part of the book has covered some of the most frequently used tools and concepts of time series regression. Many other tools for analyzing economic time series have been developed for specific applications. If you are interested in learning more about economic forecasting, see the introductory textbooks by Enders (1995) and Diebold (2007). For an advanced treatment of econometrics with time series data, see Hamilton (1994).

FIGURE 16.4 Daily Percentage Changes in the NYSE Index and GARCH(1,1) Bands



The GARCH(1,1) bands, which are $\pm\hat{\sigma}_t$, where $\hat{\sigma}_t$ is computed using Equation (16.33), are narrow when the conditional variance is small and wide when it is large. The conditional volatility of stock price changes varies considerably over the 1990–2005 period.

Summary

1. Vector autoregressions model a "vector" of k time series variables as each depends on its own lags and the lags of the $k - 1$ other series. The forecasts of each of the time series produced by a VAR are mutually consistent, in the sense that they are based on the same information.
2. Forecasts two or more periods ahead can be computed either by iterating forward a one-step-ahead model (an AR or a VAR) or by estimating a multiperiod-ahead regression.

3. Two series that share a common stochastic trend are cointegrated; that is, Y_t and X_t are cointegrated if Y_t and X_t are $I(1)$ but $Y_t - \theta X_t$ is $I(0)$. If Y_t and X_t are cointegrated, the error correction term $Y_t - \theta X_t$ can help to predict ΔY_t and/or ΔX_t. A vector error correction model is a VAR model of ΔY_t and ΔX_t, augmented to include the lagged error correction term.
4. Volatility clustering—when the variance of a series is high in some periods and low in others—is common in economic time series, especially financial time series.
5. The ARCH model of volatility clustering expresses the conditional variance of the regression error as a function of recent squared regression errors. The GARCH model augments the ARCH model to include lagged conditional variances as well. Estimated ARCH and GARCH models produce forecast intervals with widths that depend on the volatility of the most recent regression residuals.

Key Terms

vector autoregression (VAR) (632)
iterated multiperiod AR forecast (639)
iterated multiperiod VAR forecast (639)
direct multiperiod forecast (641)
integrated of order d, $I(d)$ (643)
second difference (643)
integrated of order zero [$I(0)$], one [$I(1)$], or two [$I(2)$] (643)
order of integration (643)
DF-GLS test (644)
common trend (649)
cointegration (649)
cointegrating coefficient (650)
error correction term (651)
vector error correction model (651)
EG-ADF test (653)
dynamic OLS (DOLS) estimator (654)
volatility clustering (660)
autoregressive conditional heteroskedasticity (ARCH) (660)
generalized ARCH (GARCH) (660)

Review the Concepts

16.1 A macroeconomist wants to construct forecasts for the following macroeconomic variables: GDP, consumption, investment, government purchases, exports, imports, short-term interest rates, long-term interest rates, and the rate of price inflation. He has quarterly time series for each of these variables from 1970 to 2010. Should he estimate a VAR for these variables and use this for forecasting? Why or why not? Can you suggest an alternative approach?

16.2 Suppose that Y_t follows a stationary AR(1) model with $\beta_0 = 0$ and $\beta_1 = 0.7$. If $Y_t = 5$, what is your forecast of Y_{t+2} (that is, what is $Y_{t+2|t}$)? What is $Y_{t+h|t}$ for $h = 30$? Does this forecast for $h = 30$ seem reasonable to you?

16.3 A version of the permanent income theory of consumption implies that the logarithm of real GDP (Y) and the logarithm of real consumption (C) are cointegrated with a cointegrating coefficient equal to 1. Explain how you would investigate this implication by (a) plotting the data and (b) using a statistical test.

16.4 Consider the ARCH model, $\sigma_t^2 = 1.0 + 0.8u_{t-1}^2$. Explain why this will lead to volatility clustering. (*Hint:* What happens when u_{t-1}^2 is unusually large?)

16.5 The DF-GLS test for a unit root has higher power than the Dickey–Fuller test. Why should you use a more powerful test?

Exercises

16.1 Suppose that Y_t follows a stationary AR(1) model, $Y_t = \beta_0 + \beta_1 Y_{t-1} + u_t$.

a. Show that the h-period ahead forecast of Y_t is given by $Y_{t+h|t} = \mu_Y + \beta_1^h(Y_t - \mu_Y)$, where $\mu_Y = \beta_0/(1 - \beta_1)$.

b. Suppose that X_t is related to Y_t by $X_t = \sum_{i=0}^{\infty}\delta^i Y_{t+i|t}$, where $|\delta| < 1$. Show that $X_t = \mu_Y/(1 - \delta) + (Y_t - \mu_Y)/(1 - \beta_1\delta)$.

16.2 One version of the expectations theory of the term structure of interest rates holds that a long-term rate equals the average of the expected values of short-term interest rates into the future, plus a term premium that is $I(0)$. Specifically, let Rk_t denote a k-period interest rate, let $R1_t$ denote a one-period interest rate, and let e_t denote an $I(0)$ term premium. Then $Rk_t = \frac{1}{k}\sum_{i=0}^{k-1} R1_{t+i|t} + e_t$, where $R1_{t+i|t}$ is the forecast made at date t of the value of $R1$ at date $t + i$. Suppose that $R1_t$ follows a random walk so that $R1_t = R1_{t-1} + u_t$.

a. Show that $Rk_t = R1_t + e_t$.

b. Show that Rk_t and $R1_t$ are cointegrated. What is the cointegrating coefficient?

c. Now suppose that $\Delta R1_t = 0.5\Delta R1_{t-1} + u_t$. How does your answer to (b) change?

d. Now suppose that $R1_t = 0.5R1_{t-1} + u_t$. How does your answer to (b) change?

16.3 Suppose that u_t follows the ARCH process, $\sigma_t^2 = 1.0 + 0.5u_{t-1}^2$.

a. Let $E(u_t^2) = \text{var}(u_t)$ be the unconditional variance of u_t. Show that $\text{var}(u_t) = 2$. (*Hint:* Use the law of iterated expectations $E(u_t^2) = E[E(u_t^2|u_{t-1})]$.)

b. Suppose that the distribution of u_t conditional on lagged values of u_t is $N(0, \sigma_t^2)$. If $u_{t-1} = 0.2$, what is $\Pr(-3 \le u_t \le 3)$? If $u_{t-1} = 2.0$, what is $\Pr(-3 \le u_t \le 3)$?

16.4 Suppose that Y_t follows the AR(p) model $Y_t = \beta_0 + \beta_1 Y_{t-1} + \cdots + \beta_p Y_{t-p} + u_t$, where $E(u_t|Y_{t-1}, Y_{t-2}, \ldots) = 0$. Let $Y_{t+h|t} = E(Y_{t+h}|Y_t, Y_{t-1}, \ldots)$. Show that $Y_{t+h|t} = \beta_0 + \beta_1 Y_{t-1+h|t} + \cdots + \beta_p Y_{t-p+h|t}$ for $h > p$.

16.5 Verify Equation (16.20). [*Hint:* Use $\sum_{t=1}^T Y_t^2 = \sum_{t=1}^T (Y_{t-1} + \Delta Y_t)^2$ to show that $\sum_{t=1}^T Y_t^2 = \sum_{t=1}^T Y_{t-1}^2 + 2\sum_{t=1}^T Y_{t-1}\Delta Y_t + \sum_{t=1}^T \Delta Y_t^2$ and solve for $\sum_{t=1}^T Y_{t-1}\Delta Y_t$.]

16.6 A regression of Y_t onto current, past, and future values of X_t yields

$$Y_t = 3.0 + 1.7X_{t+1} + 0.8X_t - 0.2X_{t-1} + u_t.$$

a. Rearrange the regression so that it has the form shown in Equation (16.25). What are the values of θ, δ_{-1}, δ_0, and δ_1?

b. i. Suppose that X_t is $I(1)$ and u_t is $I(1)$. Are Y and X cointegrated?

ii. Suppose that X_t is $I(0)$ and u_t is $I(1)$. Are Y and X cointegrated?

iii. Suppose that X_t is $I(1)$ and u_t is $I(0)$. Are Y and X cointegrated?

16.7 Suppose that $\Delta Y_t = u_t$, where u_t is i.i.d. $N(0, 1)$, and consider the regression $Y_t = \beta X_t + error$, where $X_t = \Delta Y_{t+1}$ and *error* is the regression error. Show that $\hat{\beta} \xrightarrow{d} \frac{1}{2}(\chi_1^2 - 1)$. [*Hint:* Analyze the numerator of $\hat{\beta}$ using analysis like that in Equation (16.21). Analyze the denominator using the law of large numbers.]

16.8 Consider the following two-variable VAR model with one lag and no intercept:

$$Y_t = \beta_{11} Y_{t-1} + \gamma_{11} X_{t-1} + u_{1t}$$

$$X_t = \beta_{21} Y_{t-1} + \gamma_{21} X_{t-1} + u_{2t}.$$

a. Show that the iterated two-period-ahead forecast for Y can be written as $Y_{t|t-2} = \delta_1 Y_{t-2} + \delta_2 X_{t-2}$ and derive values for δ_1 and δ_2 in terms of the coefficients in the VAR.

b. In light of your answer to (a), do iterated multiperiod forecasts differ from direct multiperiod forecasts? Explain.

16.9 a. Suppose that $E(u_t|u_{t-1}, u_{t-2}, \ldots) = 0$, that $\text{var}(u_t|u_{t-1}, u_{t-2}, \ldots)$ follows the ARCH(1) model $\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2$, and that the process for u_t is stationary. Show that $\text{var}(u_t) = \alpha_0/(1 - \alpha_1)$. (*Hint:* Use the law of iterated expectations $E(u_t^2) = E[E(u_t^2|u_{t-1})]$.)

b. Extend the result in (a) to the ARCH(p) model.

c. Show that $\sum_{i=1}^p \alpha_i < 1$ for a stationary ARCH(p) model.

d. Extend the result in (a) to the GARCH(1,1) model.

e. Show that $\alpha_1 + \phi_1 < 1$ for a stationary GARCH(1,1) model.

16.10 Consider the cointegrated model $Y_t = \theta X_t + v_{1t}$ and $X_t = X_{t-1} + v_{2t}$, where v_{1t} and v_{2t} are mean zero serially uncorrelated random variables with $E(v_{1t}v_{2j}) = 0$ for all t and j. Derive the vector error correction model [Equations (16.22) and (16.23)] for X and Y.

Empirical Exercises

These exercises are based on data series in the data files **USMacro_Quarterly** and **USMacro_Monthly** described in the Empirical Exercises in Chapters 14 and 15. Let $Y_t = \ln(GDP_t)$, R_t denote the 3-month Treasury bill rate, and π_t^{CPI} and π_t^{PCE} denote the inflation rates from the CPI and Personal Consumption Expenditures (PCE) Deflator, respectively.

E16.1 Using quarterly data from 1955:1 through 2009:4, estimate a VAR(4) (a VAR with four lags) for ΔY_t and ΔR_t.

a. Does ΔR Granger-cause ΔY? Does ΔY Granger-cause ΔR?

b. Should the VAR include more than four lags?

E16.2 In this exercise you will compute pseudo out-of-sample two-quarter-ahead forecasts for ΔY beginning in 1989:4 through the end of the sample. (That is, you will compute $\Delta Y_{1990:2/1989:4}$, $\Delta Y_{1990:3/1990:1}$, and so forth.)

a. Construct iterated two-quarter-ahead pseudo out-of-sample forecasts using an AR(1) model.

b. Construct iterated two-quarter-ahead pseudo out-of-sample forecasts using a VAR(4) model for ΔY and ΔR.

c. Construct iterated two-quarter-ahead pseudo out-of-sample forecasts using the naive forecast $\Delta Y_{t+2/t} = (\Delta Y_t + \Delta Y_{t-1} + \Delta Y_{t-2} + \Delta Y_{t-3})/4$.

d. Which model has the smallest root mean squared forecast error?

E16.3 Use the DF-GLS test to test for a unit autoregressive root for Y_t. As an alternative, suppose that Y_t is stationary around a deterministic trend. Compare the results to the results obtained in Empirical Exercise 14.3.

E16.4 In Empirical Exercise 15.2, you studied the behavior of $\pi_t^{CPI} - \pi_t^{PCE}$ over the sample period 1970:1 through 2009:12. That analysis was predicated on the assumption that $\pi_t^{CPI} - \pi_t^{PCE}$ is $I(0)$.

a. Test for a unit root in the autoregression for $\pi_t^{CPI} - \pi_t^{PCE}$. Carry out the test using the ADF test that includes a constant and 12 lags of the first difference of $\pi_t^{CPI} - \pi_t^{PCE}$. Also carry out the test using the DF-GLS procedure.

b. Test for a unit root in the autoregression for π_t^{CPI} and in the autoregression for π_t^{PCE}. As in (a), use both the ADF and DF-GLS tests including a constant and 12 lagged first differences.

c. What do the results from (a) and (b) say about cointegration between these two inflation rates? What is the value of the cointegrating coefficient (θ) implied by your answers to (a) and (b)?

d. Suppose that you did not know that the cointegrating coefficient was $\theta = 1$. How would you test for cointegration? Carry out the test. How would you estimate θ? Estimate the value of θ using the DOLS regression of π_t^{CPI} onto π_t^{PCE} and six leads and lags of $\Delta\pi_t^{PCE}$. Is the estimated value of θ close to 1?

E16.5 **a.** Using data on ΔY (the growth rate in GDP) from 1955:1 to 2009:4, estimate an AR(1) model with GARCH(1,1) errors.

b. Plot the residuals from the AR(1) model along with $\pm\hat{\sigma}_t$ bands as in Figure 16.4.

c. Some macroeconomists have claimed that there was a sharp drop in the variability of ΔY around 1983, which they call the "Great Moderation." Is this Great Moderation evident in the plot that you formed in (b)?

APPENDIX

16.1 U.S. Financial Data Used in Chapter 16

The interest rates on 3-month U.S. Treasury bills and on 1-year U.S. Treasury bonds are the monthly average of their daily rates, converted to an annual basis, as reported by the Board of Governors of the U.S. Federal Reserve. The quarterly data used in this chapter are the monthly average interest rates for the final month in the quarter.

CHAPTER 17 The Theory of Linear Regression with One Regressor

Why should an applied econometrician bother learning any econometric theory? There are several reasons. Learning econometric theory turns your statistical software from a "black box" into a flexible toolkit from which you are able to select the right tool for the job at hand. Understanding econometric theory helps you appreciate why these tools work and what assumptions are required for each tool to work properly. Perhaps most importantly, knowing econometric theory helps you recognize when a tool will *not* work well in an application and when you should look for a different econometric approach.

This chapter provides an introduction to the econometric theory of linear regression with a single regressor. This introduction is intended to supplement—not replace—the material in Chapters 4 and 5, which should be read first.

This chapter extends Chapters 4 and 5 in two ways.

First, it provides a mathematical treatment of the sampling distribution of the OLS estimator and *t*-statistic, both in large samples under the three least squares assumptions of Key Concept 4.3 and in finite samples under the two additional assumptions of homoskedasticity and normal errors. These five extended least squares assumptions are laid out in Section 17.1. Sections 17.2 and 17.3, augmented by Appendix 17.2, develop mathematically the large-sample normal distributions of the OLS estimator and *t*-statistic under the first three assumptions (the least squares assumptions of Key Concept 4.3). Section 17.4 derives the exact distributions of the OLS estimator and *t*-statistic under the two additional assumptions of homoskedasticity and normally distributed errors.

Second, this chapter extends Chapters 4 and 5 by providing an alternative method for handling heteroskedasticity. The approach of Chapters 4 and 5 is to use heteroskedasticity-robust standard errors to ensure that statistical inference is valid even if the errors are heteroskedastic. This method comes with a cost, however: If the errors are heteroskedastic, then in theory a more efficient estimator than OLS is available. This estimator, called weighted least squares, is presented in Section 17.5. Weighted least squares requires a great deal of prior knowledge about the precise nature of the heteroskedasticity—that is, about the conditional variance of u given X. When such knowledge is available, weighted least squares improves upon OLS. In