CHAPTER 17 Additional Topics in Time Series Regression

This chapter takes up some further topics in time series regression, starting with forecasting. Chapter 15 considered forecasting a single variable. In practice, however, you might want to forecast two or more variables, such as the growth rate of gross domestic product (GDP) and the rate of inflation. Section 17.1 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 15 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 multi-period forecasts are discussed in Section 17.2.

Sections 17.3 and 17.4 return to the topic of Section 15.6, stochastic trends. Section 17.3 introduces additional models of stochastic trends. Section 17.4 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 17.5 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 or forecasting the price of derivative assets that depend on this risk.

Section 17.6 takes up the challenge of forecasting when there are many predictors, as is the case for macroeconomic data in developed economies. This section draws on material introduced in Chapter 14 and focuses on one commonly used method for forecasting with large data sets, which uses principal components analysis to reduce the information in a large time series data set to a small number of time series. The framework for doing so is the dynamic factor model, which also can be used for purposes other than forecasting.

17.1 Vector Autoregressions

Chapter 15 focused on forecasting the growth rate of GDP, but in reality, economic forecasters are in the business of forecasting other key macroeconomic variables as well, such as the rate of inflation, the unemployment rate, and interest rates. One

KEY CONCEPT 17.1

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,

$$Y_t = \beta_{10} + \beta_{11}Y_{t-1} + \cdots + \beta_{1p}Y_{t-p} + \gamma_{11}X_{t-1} + \cdots + \gamma_{1p}X_{t-p} + u_{1t} \tag{17.1}$$

$$X_t = \beta_{20} + \beta_{21}Y_{t-1} + \cdots + \beta_{2p}Y_{t-p} + \gamma_{21}X_{t-1} + \cdots + \gamma_{2p}X_{t-p} + u_{2t}, \tag{17.2}$$

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 15.6 applied to each equation. The coefficients of a VAR are estimated by estimating each equation by OLS.

approach is to develop a separate forecasting model for each variable, using the methods of Section 15.4. 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 ordinary least squares (OLS).

VARs are summarized in Key Concept 17.1.

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 (17.1) and (17.2), 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 0 in these two equations. The null hypothesis that these coefficients are 0 is

$$H_0: \beta_{1p} = 0,\ \beta_{2p} = 0,\ \gamma_{1p} = 0, \text{ and } \gamma_{2p} = 0. \tag{17.3}$$

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 5 variables and 4 lags will have 21 coefficients (4 lags each of 5 variables, plus the intercept) in each of the 5 equations, for a total of 105 coefficients! As discussed in Section 14.2, estimating all these coefficients by OLS increases the amount of estimation error entering a forecast, which can result in deterioration of the accuracy of the forecast as measured by the mean squared forecast error (MSFE). If the VAR coefficients are estimated by OLS, the number of coefficients therefore should be small relative to the sample size, so the number of VAR variables should be few.

In this section, we consider small VARs with coefficients estimated by OLS. Because a small VAR has only a handful of variables, those variables should be chosen with care. One guideline is to make sure 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 15) and economic theory that the growth rate of GDP, the term spread, and the rate of inflation are related to one another, suggesting that these variables could help forecast one another in a VAR. Including an unrelated variable in a VAR, however, introduces estimation error without adding predictive content, thereby reducing forecast accuracy.

An alternative approach is to use many variables but to use methods other than OLS. We take up forecasting with many predictors in Section 17.6.

Determining lag lengths in VARs. 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 15.5. To define this information criterion, we need to adopt matrix notation (reviewed in Appendix 19.1). 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 Bayes information criterion (BIC) for the VAR is

$$\text{BIC}(p) = \ln[\det(\hat{\Sigma}_u)] + k(kp + 1)\frac{\ln(T)}{T}, \tag{17.4}$$

where $\det(\hat{\Sigma}_u)$ is the determinant of the matrix $\hat{\Sigma}_u$. The Akaike information criterion (AIC) is computed using Equation (17.4), modified by replacing the term $\ln(T)$ with 2.

The expression for the BIC for the k equations in the VAR in Equation (17.4) extends the expression for a single equation given in Section 15.5. When there is a single equation, the first term simplifies to $\ln[SSR(p)/T]$. The second term in Equation (17.4) 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). (See the box "Nobel Laureates in Time Series Econometrics.") 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 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 a graduate textbook treatment of structural VAR modeling, see Kilian and Lütkepohl (2017).

A VAR Model of the Growth Rate of GDP and the Term Spread

As an illustration, consider a two-variable VAR for the growth rate of GDP, $GDPGR_t$, and the term spread, $TSpread_t$. The VAR for $GDPGR_t$ and $TSpread_t$ consists of two equations: one in which $GDPGR_t$ is the dependent variable and one in which $TSpread_t$ is the dependent variable. The regressors in both equations are lagged values of $GDPGR_t$ and $TSpread_t$. Because of the apparent break in the relation in the early 1980s found in Section 15.7 using the Quandt likelihood ratio (QLR) test, the VAR is estimated using data from 1981:Q1 to 2017:Q3.

The first equation of the VAR is the GDP growth rate equation:

$$\begin{aligned}\widehat{GDPGR}_t = \underset{(0.50)}{0.54} &+ \underset{(0.11)}{0.29}\,GDPGR_{t-1} + \underset{(0.08)}{0.20}\,GDPGR_{t-2} \\ &- \underset{(0.35)}{0.86}\,TSpread_{t-1} + \underset{(0.39)}{1.18}\,TSpread_{t-2}.\end{aligned} \tag{17.5}$$

The adjusted R^2 is $\overline{R}^2 = 0.27$.

The second equation of the VAR is the term spread equation, in which the regressors are the same as in the $GDPGR$ equation but the dependent variable is the term spread:

$$\begin{aligned}\widehat{TSpread}_t = \underset{(0.12)}{0.44} &+ \underset{(0.02)}{0.01}\,GDPGR_{t-1} - \underset{(0.03)}{0.05}\,GDPGR_{t-2} \\ &+ \underset{(0.10)}{1.06}\,TSpread_{t-1} - \underset{(0.11)}{0.22}\,TSpread_{t-2}.\end{aligned} \tag{17.6}$$

The adjusted R^2 is $\overline{R}^2 = 0.82$.

Equations (17.5) and (17.6), taken together, are a VAR(2) model of the growth rate of GDP, $GDPGR_t$, and the term spread, $TSpread_t$.

These VAR equations can be used to perform tests of predictability. The F-statistic testing the null hypothesis that the coefficients on $TSpread_{t-1}$ and $TSpread_{t-2}$ are 0 in the GDP growth rate equation [Equation (17.5)] is 5.60, which has a p-value less than 0.001. Thus the null hypothesis is rejected, so we can conclude that the term spread is a useful predictor of the growth rate of GDP, given lags in the growth rate of GDP. The F-statistic testing the hypothesis that the coefficients on the two lags of $GDPGR_t$ are zero in the term spread equation [Equation (17.6)] is 3.22, which has a p-value of 0.04. Thus the growth rate of GDP helps predict the term spread at the 5% significance level.

Forecasts of the growth rate of GDP and the term spread one period ahead are obtained exactly as discussed in Section 15.4. The forecast of the growth rate of GDP for 2017:Q4, based on Equation (17.5), is $\widehat{GDP}_{2017:Q4|2017:Q3} = 2.8\%$. A similar calculation using Equation (17.6) gives a forecast of the term spread for 2017:Q4, based on data through 2017:Q3, of $\widehat{TSpread}_{2017:Q4|2017:Q3} = 1.3$ percentage points. The actual values for 2017:Q4 are $GDPGR_{2017:Q4} = 2.5\%$ and $TSpread_{2017:Q4} = 1.2$ percentage points.

17.2 Multi-period 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 multi-period forecasts, which are also called multi-step forecasts. The first 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 multi-period 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 Multi-period 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 $GDPGR$ [Equation (15.9)]:

$$\underset{}{\widehat{GDPGR}_t} = \underset{(0.32)}{1.95} + \underset{(0.07)}{0.34}\, GDPGR_{t-1}. \tag{17.7}$$

The first step in computing the two-quarter ahead forecast of $GDPGR_{2018:Q1}$ based on Equation (17.7) and using data through 2017:Q3 is to compute the one-quarter ahead forecast of $GDPGR_{2017:Q4}$ based on data through 2017:Q3: $\widehat{GDPGR}_{2017:Q4|2017:Q3} = 1.95 + 0.34\, GDPGR_{2017:Q3} = 1.95 + 0.34 \times 3.11 = 3.0$. The second step is to substitute this forecast into Equation (17.7), so that $\widehat{GDPGR}_{2018:Q1|2017:Q3} = 1.95 + 0.34\, GDPGR_{2017:Q4|2017:Q3} = 1.95 + 0.34 \times 3.0 = 3.0$. Thus, based on information through the third quarter of 2017, this forecast states that the growth rate of GDP will be 3.0% in the first quarter of 2018.

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 the growth rate of GDP based on the AR(2) model from Section 15.3 [Equation (15.11)]:

$$\underset{}{\widehat{GDPGR}_t} = \underset{(0.37)}{1.60} + \underset{(0.08)}{0.28}\,GDPGR_{t-1} + \underset{(0.08)}{0.18}\,GDPGR_{t-2}. \tag{17.8}$$

The forecast of $GDPGR_{2017:Q4}$ based on data through 2017:Q3 using this AR(2), computed in Section 15.3, is $\widehat{GDPGR}_{2017:Q4|2017:Q3} = 3.0$. Thus the two-quarter ahead iterated forecast based on the AR(2) is $\widehat{GDPGR}_{2018:Q1|2017:Q3} = 1.60 + 0.28\,\widehat{GDPGR}_{2017:Q4|2017:Q3} + 0.18\,GDPGR_{2017:Q3} = 1.60 + 0.28 \times 3.0 + 0.18 \times 3.1 = 3.0$. According to this iterated AR(2) forecast, based on data through the third quarter of 2017, the growth rate of GDP is predicted to be 3.0% in the first quarter of 2018.

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 growth rate of GDP in period $T + 2$ using a VAR with the variables $GDPGR_t$ and $TSpread_t$, one must forecast both $GDPGR_{T+1}$ and $TSpread_{T+1}$, using data through period T as an intermediate step in forecasting $GDPGR_{T+2}$. More generally, to compute multi-period 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 $GDPGR_{2018:Q1}$ based on data through 2017:Q3, using the VAR(2) for $GDPGR_t$ and $TSpread_t$ in Section 17.1 [Equations (17.5) and (17.6)]. The first step is to compute the one-quarter ahead forecasts $\widehat{GDPGR}_{2017:Q4|2017:Q3}$ and $\widehat{TSpread}_{2017:Q4|2017:Q3}$ from that VAR. These one-period ahead forecasts were computed in Section 17.1 based on Equations (17.5) and (17.6). The forecasts were $\widehat{GDPGR}_{2017:Q4|2017:Q3} = 2.8$ and $\widehat{TSpread}_{2017:Q4|2017:Q3} = 1.3$. In the second step, these forecasts are substituted into Equations (17.5) and (17.6) to produce the two-quarter ahead forecast:

$$\begin{aligned}\widehat{GDPGR}_{2018:Q1|2017:Q3} &= 0.54 + 0.29\,\widehat{GDPGR}_{2017:Q4|2017:Q3} + 0.20GDPGR_{2017:Q3} \\ &\quad -0.86\,\widehat{TSpread}_{2017:Q4|2017:Q3} + 1.28TSpread_{2017:Q3} \\ &= 0.54 + 0.29 \times 2.8 + 0.20 \times 3.1 \\ &\quad -0.86 \times 1.3 + 1.28 \times 1.2 = 2.4.\end{aligned} \tag{17.9}$$

Thus the iterated VAR(2) forecast, based on data through the third quarter of 2017, is that the growth rate of GDP will be 2.4% in the first quarter of 2018.

Iterated multi-period forecasts are summarized in Key Concept 17.2.

KEY CONCEPT 17.2

Iterated Multi-period Forecasts

The **iterated multi-period AR forecast** is computed in steps: First compute the one-period ahead forecast, and 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

$$\hat{Y}_{T+2|T} = \hat{\beta}_0 + \hat{\beta}_1\hat{Y}_{T+1|T} + \hat{\beta}_2 Y_T + \hat{\beta}_3 Y_{T+1} + \cdots + \hat{\beta}_p Y_{T-p+2} \qquad (17.10)$$

$$\hat{Y}_{T+3|T} = \hat{\beta}_0 + \hat{\beta}_1\hat{Y}_{T+2|T} + \hat{\beta}_2\hat{Y}_{T+1|T} + \hat{\beta}_3 Y_T + \cdots + \hat{\beta}_p Y_{T-p+3}, \qquad (17.11)$$

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 multi-period 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 17.1, is

$$\begin{aligned}\hat{Y}_{T+2|T} = {}& \hat{\beta}_{10} + \hat{\beta}_{11}\hat{Y}_{T+1|T} + \hat{\beta}_{12} Y_T + \hat{\beta}_{13} Y_{T-1} + \cdots + \hat{\beta}_{1p} Y_{T-p+2} \\ & + \hat{\gamma}_{11}\hat{X}_{T+1|T} + \hat{\gamma}_{12} X_T + \hat{\gamma}_{13} X_{T-1} + \cdots + \hat{\gamma}_{1p} X_{T-p+2},\end{aligned} \qquad (17.12)$$

where the coefficients in Equation (17.12) are the OLS estimates of the VAR coefficients. Iterating produces forecasts further into the future.

Direct Multi-period Forecasts

Direct multi-period forecasts are computed without iterating by using a single regression, in which the dependent variable is the multi-period 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 multi-period forecast.

The direct multi-period 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 $GDPGR_t$ two quarters ahead using two lags each of $GDPGR_{t-2}$ and $TSpread_{t-2}$ is computed by first estimating the regression:

$$\begin{aligned}\widehat{GDPGR}_{t|t-2} = \underset{(0.63)}{0.56} + \underset{(0.07)}{0.31}GDPGR_{t-2} + \underset{(0.09)}{0.04}GDPGR_{t-3} \\ + \underset{(0.46)}{0.56}TSpread_{t-2} + \underset{(0.45)}{0.04}TSpread_{t-3}.\end{aligned} \tag{17.13}$$

The two-quarter ahead forecast of the growth rate of GDP in 2018:Q1 based on data through 2017:Q3 is computed by substituting the values of $GDPGR_{2017:Q3}$, $GDPGR_{2017:Q2}$, $TSpread_{2017:Q3}$, and $TSpread_{2017:Q2}$ into Equation (17.13); this yields

$$\begin{aligned}\widehat{GDPGR}_{2018:Q1|2017:Q3} = 0.56 + 0.31GDPGR_{2017:Q3} + 0.04GDPGR_{2017:Q2} \\ + 0.56TSpread_{2017:Q3} + 0.04TSpread_{2017:Q2} = 2.4.\end{aligned} \tag{17.14}$$

The three-quarter ahead direct forecast of $GDPGR_{T+3}$ is computed by lagging all the regressors in Equation (17.13) by one additional quarter, estimating that regression, and then computing the forecast. The h-quarter ahead direct forecast of $GDPGR_{T+h}$ is computed by using $GPDGR_t$ as the dependent variable and the regressors $GPDGR_{t-h}$ and $TSpread_{t-h}$ plus additional lags of $GPDGR_{t-h}$ and $TSpread_{t-h}$, as desired.

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

As discussed in Section 16.4, 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 multi-period regressions. The standard errors reported in Equation (17.13) for direct multi-period regressions therefore are Newey–West HAC standard errors, where the truncation parameter m is set according to Equation (16.17); for these data (for which $T = 147$), Equation (16.17) yields $m = 4$. 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 (16.17) are appropriate for multi-period regressions with long forecast horizons.

Direct multi-period forecasts are summarized in Key Concept 17.3.

KEY CONCEPT 17.3

Direct Multi-period Forecasts

The **direct multi-period 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

$$Y_t = \delta_0 + \delta_1 Y_{t-h} + \cdots + \delta_p Y_{t-p-h+1} + \delta_{p+1} X_{t-h} + \cdots + \delta_{2p} X_{t-p-h+1} + u_t \tag{17.15}$$

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 multi-period 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 multi-period 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 multi-period 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 are specified incorrectly, perhaps because of neglected nonlinear terms. If the one-step ahead model is specified incorrectly, then, in general, the iterated multi-period 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.

17.3 Orders of Integration and the Nonnormality of Unit Root Test Statistics

This section extends the treatment of stochastic trends in Section 15.6 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. Next we discuss the reason for the nonnormal distribution of the ADF test for a unit root.

Other Models of Trends and Orders of Integration

Recall that the random walk model for a trend, introduced in Section 15.6, 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

$$Y_t = \beta_0 + Y_{t-1} + u_t, \tag{17.16}$$

where u_t is serially uncorrelated. Also recall from Section 15.6 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, that vary less from one period to the next—than is implied by Equation (17.16). A different model is needed to describe the trends of such series.

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

$$\Delta Y_t = \beta_0 + \Delta Y_{t-1} + u_t, \tag{17.17}$$

where u_t is serially uncorrelated. Thus, if Y_t follows Equation (17.17), 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 (17.17), then its second difference is stationary. If a series has a trend of the form in Equation (17.17), 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 (17.17) 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 17.4.

KEY CONCEPT 17.4

Orders of Integration, Differencing, and Stationarity

- 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 $\boldsymbol{I(d)}$—then Y_t must be differenced d times to eliminate its stochastic trend; that is, $\Delta^d Y_t$ is stationary.

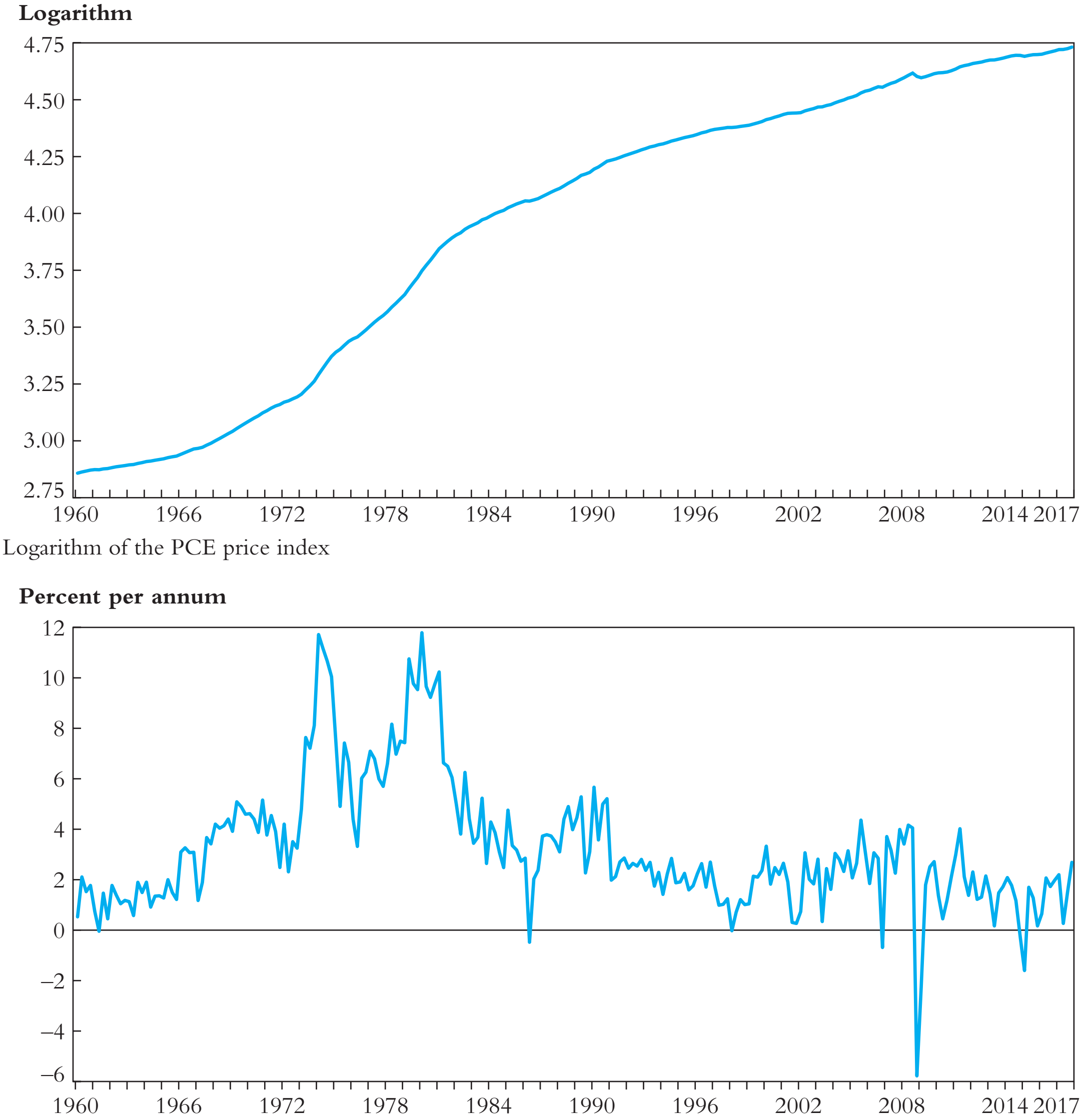
How to test whether a series is I(2) or I(1). If Y_t is $I(2)$, then ΔY_t is $I(1)$, so ΔY_t has an autoregressive root that equals 1. If, however, Y_t is $I(1)$, then ΔY_t is stationary. Thus the null hypothesis that Y_t is $I(2)$ can be tested against the alternative hypothesis that Y_t is $I(1)$ by 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. The rate of inflation is the growth rate of the price level. Recall from Section 15.2 that the growth rate of a time series X_t can be computed as the first difference of the logarithm of X_t; that is, $\Delta\ln(X_t)$ is the growth rate of X_t (expressed as fraction). If P_t is a time series for the price level measured quarterly, then $\Delta\ln(P_t)$ is its growth rate, and $Infl_t = 400 \times \Delta\ln(P_t)$ is the quarterly rate of inflation, measured in percentage points at an annual rate. As in the expression for the growth of GDP, $GDPGR$ in Equation (15.1), the factor 400 arises from converting fractional changes to percentage changes (multiplying by 100) and converting quarterly percentages to an annual rate (multiplying by 4).

In Empirical Exercise 15.1, you analyzed the inflation rate, $Infl_t$, computed using the price index for personal consumption expenditures in the United States as P_t. In that exercise, you 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 $\Delta Infl_t$ is stationary. But treating inflation as $I(1)$ is equivalent to treating $\Delta\ln(P_t)$ as $I(1)$, and this in turn is equivalent to treating the logarithm of the price level, $\ln(P_t)$, as $I(2)$.

The logarithm of the price level and the rate of inflation are plotted in Figure 17.1. The long-run trend of the logarithm of the price level (Figure 17.1a) varies more smoothly than the long-run trend in the rate of inflation (Figure 17.1b). The smooth trend in the logarithm of the price level is typical of $I(2)$ series.

FIGURE 17.1 The Logarithm of the Price Level and the Inflation Rate in the United States, 1960–2017



(a) Logarithm of the PCE price index

(b) Inflation

The trend in the logarithm of prices (Figure 17.1a) is much smoother than the trend in inflation (Figure 17.1b).

Why Do Unit Root Tests Have Nonnormal Distributions?

In Section 15.7, it was stressed that the large-sample normal distribution on 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 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 Equation (15.32), 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

$$T\hat{\delta} = \frac{\frac{1}{T}\sum_{t=1}^{T} Y_{t-1}\Delta Y_t}{\frac{1}{T^2}\sum_{t=1}^{T} Y_{t-1}^2}. \tag{17.19}$$

Consider the numerator in Equation (17.19). Under the additional assumption that $Y_0 = 0$, a bit of algebra (Exercise 17.5) shows that

$$\frac{1}{T}\sum_{t=1}^{T} Y_{t-1}\Delta Y_t = \frac{1}{2}\left[\left(\frac{Y_T}{\sqrt{T}}\right)^2 - \frac{1}{T}\sum_{t=1}^{T}(\Delta Y_t)^2\right]. \tag{17.20}$$

Under the null hypothesis, $\Delta Y_t = u_t$, which is serially uncorrelated and has a finite variance, so the second term in Equation (17.20) 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 (17.20) 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 (17.20) that, under the null hypothesis, the numerator in Equation (17.19) has the limiting distribution

$$\frac{1}{T}\sum_{t=1}^{T} Y_{t-1}\Delta Y_t \xrightarrow{d} \frac{\sigma_u^2}{2}\left(\chi_1^2 - 1\right). \tag{17.21}$$

The large-sample distribution in Equation (17.21) 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 (17.19) 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 joint distribution of the numerator and denominator in Equation (17.19) 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.

17.4 Cointegration

Sometimes two or more series share the same stochastic trend. 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, the 90-day and 10-year U.S. Treasury interest rates in Figure 15.3 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, and then both fell through the 1990s. However, the difference between the long-term and short-term interest rates, the term spread shown in Figure 15.3b, does not appear to have a trend. That is, subtracting the short-term rate from the long-term 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; see the box "Nobel Laureates in Time Series Econometrics") is given in Key Concept 17.5. 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. The discussion initially focuses on the case that there are only two variables, X_t and Y_t.

Vector error correction model. 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:

$$\begin{aligned}\Delta Y_t = {} & \beta_{10} + \beta_{11}\Delta Y_{t-1} + \cdots + \beta_{1p}\Delta Y_{t-p} + \gamma_{11}\Delta X_{t-1} \\ & + \cdots + \gamma_{1p}\Delta X_{t-p} + \alpha_1(Y_{t-1} - \theta X_{t-1}) + u_{1t}\end{aligned} \tag{17.22}$$

$$\begin{aligned}\Delta X_t = {} & \beta_{20} + \beta_{21}\Delta Y_{t-1} + \cdots + \beta_{2p}\Delta Y_{t-p} + \gamma_{21}\Delta X_{t-1} \\ & + \cdots + \gamma_{2p}\Delta X_{t-p} + \alpha_2(Y_{t-1} - \theta X_{t-1}) + u_{2t}.\end{aligned} \tag{17.23}$$

The term $Y_t - \theta X_t$ is called the **error correction term**: if the two variables are far apart, by virtue of their sharing a trend, one would expect the variables to get closer together over time, so that the "error" $Y_t - \theta X_t$ will be "corrected."

The combined model in Equations (17.22) and (17.23) 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.

KEY CONCEPT 17.5

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 a common, stochastic trend. Computing the difference $Y_t - \theta X_t$ eliminates this common stochastic trend.

How Can You Tell Whether Two Variables Are Cointegrated?

There are three ways to determine whether two variables can plausibly be modeled as cointegrated: You can 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. In practice, you should use all three methods.

For example, the two interest rates in Figure 15.3 are linked together by the so-called expectations theory of the term structure of interest rates, which holds that the 10-year Treasury bond rate is the average of the sequence of expected interest rates on 3-month Treasury bills over the 10-year life of the bond. Thus, if the 3-month interest rate has a random walk stochastic trend, this theory implies that this stochastic trend is inherited by the 10-year interest rate (Exercise 17.2). Moreover, the plot of the two interest rates in Figure 15.3 shows that each of the series appears to be $I(1)$ but that the term spread appears to be $I(0)$, so it is plausible that the two series are cointegrated.

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—that 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 many cases, expert knowledge or economic theory suggests a value for θ. When θ is known, the ADF 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.

As an illustration, applying the ADF test to the term spread (the difference between the 10-year and 90-day Treasury rates) from 1962 to 2017, with an intercept and (AIC-determined) six lags, yields an ADF statistic of −4.13. This value is less

TABLE 17.1 Critical Values for the Engle–Granger ADF Statistic

Number of *X*'s in Equation (17.24)	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

than −3.43 from Table 15.4, so the null hypothesis of no cointegration (a unit root in the term spread) is rejected at the 1% significance level.

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

$$Y_t = \alpha + \theta X_t + z_t. \tag{17.24}$$

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 17.1.[1] The critical values in the first row apply when there is a single regressor in Equation (17.26), 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 (17.24) is consistent. However, in general, the OLS estimator (like the ADF test statistic, for similar reasons) 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

[1]The critical values in Table 17.1 are taken from Fuller (1976) and Phillips and Ouliaris (1990). Following a suggestion by Hansen (1992), the critical values in Table 17.1 are chosen so that they apply whether or not X_t and Y_t have drift components.

modified version of Equation (17.24) that includes past, present, and future values of the change in X_t:

$$Y_t = \beta_0 + \theta X_t + \sum_{j=-p}^{p} \delta_j \Delta X_{t-j} + u_t. \tag{17.25}$$

Thus, in Equation (17.25), 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 (17.25).

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 (17.25) 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.

As an illustration, for a DOLS regression of the 90-day Treasury rate on the 10-year Treasury rate, using the data in Figure 15.3 and $p = 4$ leads and lags, the DOLS estimate of the cointegrating coefficient is 1.02. The HAC standard error, computed using a truncation parameter of $m = 5$, is 0.05. Thus the null hypothesis that $\theta = 1$ cannot be rejected at the 10% significance level. This result, along with the finding that the term spread is stationary, is consistent with the theory of the term structure of interest rates.

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 three-month rate ($R3m$), the one-year ($R1y$) rate, and the ten-year rate ($R10y$). 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 $R10y_t - R3m_t$, and a second relationship is $R1y_t - R3m_t$. (The relationship $R10y_t - R1y_t$ 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 (17.24) is modified so that both X_{1t} and X_{2t} are regressors; the critical values for the EG-ADF test are given in Table 17.1, 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. For additional discussion of cointegration methods for multiple variables, see Hamilton (1994).

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 hypothesis), it is especially important to rely on economic theory, institutional knowledge, and common sense when estimating and using cointegrating relationships.

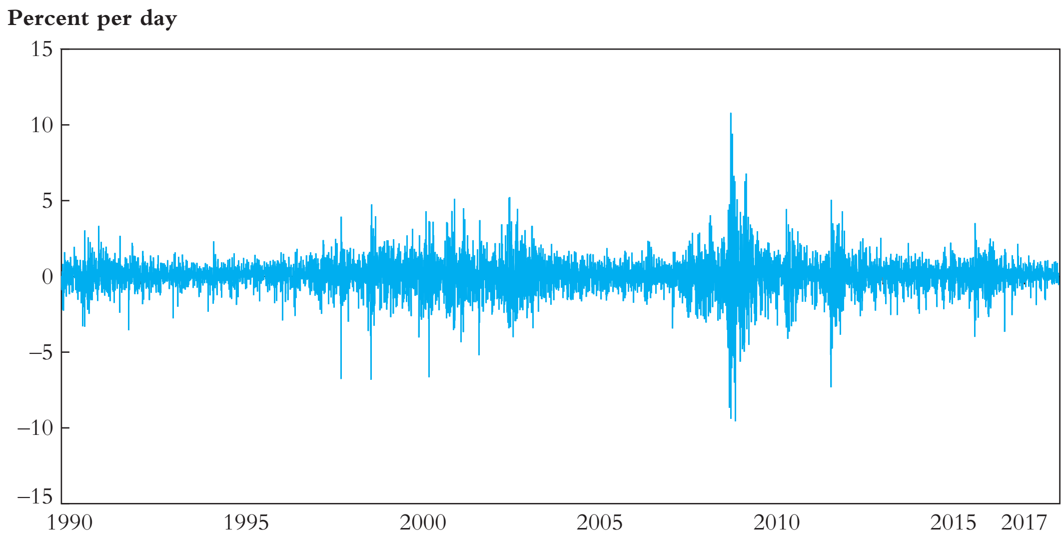
17.5 Volatility Clustering and Autoregressive Conditional Heteroskedasticity

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 Wilshire 5000 Total Market Index, shown in Figure 17.2, exhibit periods of high volatility, such as in 2001 and 2008, and other periods of low volatility, such as in 2004 and 2017. 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

FIGURE 17.2 Daily Percentage Changes in the Wilshire 5000 Total Market Index, 1990–2017



Daily percentage price changes in the Wilshire 5000 Total Market Index exhibit volatility clustering, in which there are some periods of high volatility, such as in 2008, and other periods of relative tranquility, such as in 2004.

change in the Wilshire 5000 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 price changes for some asset 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 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 can improve the accuracy of 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 using the standard error of the regression or final prediction error as discussed in Section 15.5. 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. If the forecast error changes slowly, then the pseudo out-of-sample forecast error estimate of the MSFE in Equation (15.22) can be used, but to capture more rapid changes in volatility, such as those observed in Figure 17.2, other methods must be used.

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.

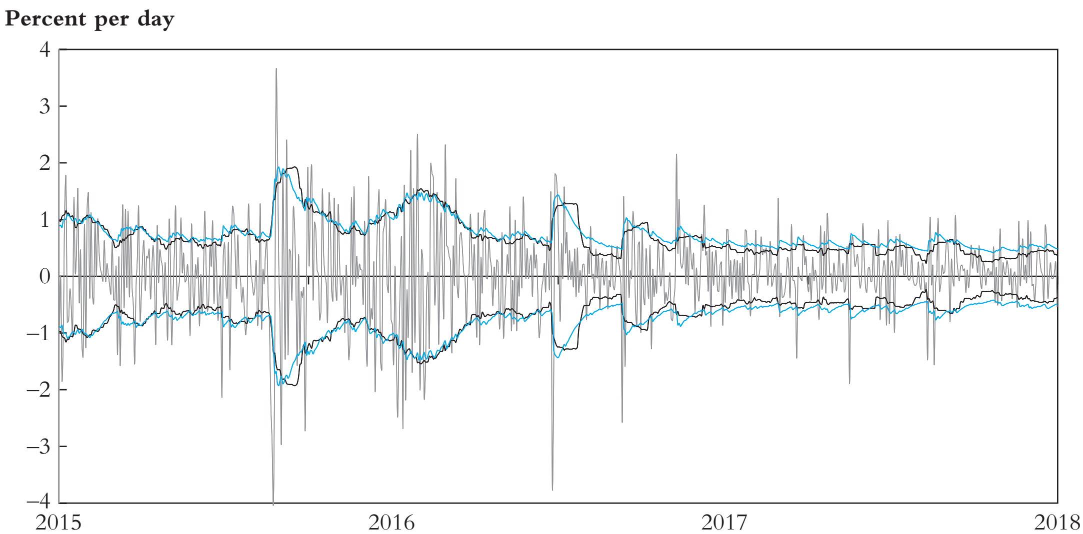
When data are observed at a high frequency, it is possible to measure volatility directly using a measure called realized volatility. When data are observed less frequently, it is possible to estimate a model of the volatility and use that to estimate current volatility. We address these two approaches in turn.

Realized Volatility

Suppose you have daily data on asset returns, like that shown in Figure 17.2. One way to estimate the volatility in a given month is to compute the sample variance of asset returns in that month. For asset returns measured at high frequency, the mean return is typically very small compared with the variation in the return, as is evident in Figure 17.2. For that reason, for asset returns, and more generally for series that can be measured at a high frequency, the volatility of the return is measured not by the sample variance but simply by its mean square. Accordingly, the h-period **realized volatility** of a variable X_t is the sample root mean square of X computed over h consecutive periods:

$$RV_t^h = \sqrt{\frac{1}{h}\sum_{s=t-h+1}^{t} X_s^2}. \tag{17.26}$$

FIGURE 17.3 Daily Percentage Changes in the Wilshire 5000 Total Market Index, 20-day Realized Volatility Bands, and GARCH(1, 1) Bands, 2015–2017



The volatility of stock price changes varies considerably over the 2015–2017 period. The volatility bands are narrow when volatility is low and wide when it is high. The 20-day realized volatility bands (black) and GARCH(1, 1) bands (dark blue) are similar to each other.

The 20-day realized volatility bands of the data in Figure 17.2 for 2015–2017 is plotted in Figure 17.3. As can be seen from the figure, the realized volatility bands provides a smooth measure of the volatility clustering evident in that figure.

In practice, realized volatility is typically computed using higher-frequency data than just daily. For example, the stock of a major company might be traded sufficiently frequently that its price can be measured at five-minute intervals. If so, these five-minute intervals can be used to compute realized volatility for a day, or even for a period of hours within a day. High-frequency realized volatility is one of the tools used in high-frequency trading.

Autoregressive Conditional Heteroskedasticity

When data are observed less frequently, an alternative is to estimate a model of the evolution of the variance over time. 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

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \gamma_1 X_{t-1} + u_t. \tag{17.27}$$

In the ARCH model, which was developed by the econometrician Robert Engle (1982; see the box "Nobel Laureates in Time Series Econometrics"), the error u_t is modeled as being normally distributed with mean 0 and variance σ_t^2, where σ_t^2 depends on past squared values of u_t. Specifically, the ARCH model of order p, denoted ARCH(p), is

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \alpha_2 u_{t-2}^2 + \cdots + \alpha_p u_{t-p}^2, \tag{17.28}$$

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 (17.27), 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 0, including higher-order ADL models, autoregressions, and time series regressions with multiple predictors.

GARCH. The 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

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \cdots + \alpha_p u_{t-p}^2 + \phi_1 \sigma_{t-1}^2 + \cdots + \phi_q \sigma_{t-q}^2, \tag{17.29}$$

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 Chapter 16, the ADL model can provide a more parsimonious model of dynamic multipliers than can 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 17.2. In such applications, the return itself is often modeled as unpredictable, so the regression in Equation (17.27) includes only 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 Wilshire 5000 daily percentage stock price changes, R_t, estimated using data on all trading days from January 2, 1990, through December 29, 2017, is

$$\underset{(0.010)}{\hat{R}_t = 0.063} \tag{17.30}$$

$$\hat{\sigma}_t^2 = \underset{(0.002)}{0.013} + \underset{(0.008)}{0.088}\, u_{t-1}^2 + \underset{(0.009)}{0.908}\sigma_{t-1}^2. \quad (17.31)$$

No lagged predictors appear in Equation (17.30) because daily Wilshire 5000 percentage 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 17.9). This sum (0.99) is large, indicating that changes in the conditional variance are persistent. Said differently, the estimated GARCH model implies that periods of high volatility in stock prices will be long lasting. This implication is consistent with the long periods of volatility clustering seen in Figure 17.2.

The estimated conditional variance at date t, $\hat{\sigma}_t^2$, can be computed using the residuals from Equation (17.30) and the coefficients in Equation (17.31). For the Wilshire 5000 returns, the GARCH(1, 1) model and the 20-day realized volatility provide quantitatively similar estimates of the time-varying standard deviation of returns. This can be seen in Figure 17.3, which focuses on the 2015–2017 sample period. During the first half of 2015, the conditional standard deviation bands are relatively tight, indicating lower levels of risk for investors holding a portfolio of stocks making up the Wilshire 5000. But in the second half of 2015 these conditional standard deviations widened, indicating greater daily stook price volatility.

For these data, the realized volatility and GARCH bands are quantitatively similar to each other. An advantage of realized volatility is that it measures the changing variance without making any modeling assumptions. An advantage of the GARCH model is that it can be used to forecast volatility; another advantage is that it can be used in applications in which the data are observed infrequently—for example, monthly or quarterly. In general, realized volatility and GARCH models provide two complementary ways to quantify volatility clustering.

17.6 Forecasting with Many Predictors Using Dynamic Factor Models and Principal Components[2]

Statistical agencies in developed economies regularly report data on hundreds or thousands of time series describing the macroeconomy. These data include detailed information from the national income and product accounts (consumption, investment, imports, exports, and government spending), multiple series on price and wage

[2]This section draws on the material in Section 14.5, which should be read first.

inflation, output and production by industry or sector, data on specific markets such as housing, and data for asset markets including interest rates and asset prices. Each of these series could potentially contain information that could improve macroeconomic forecasts. But as explained in Chapter 14, with many predictors—potentially more than the number of available time series observations—regressions estimated by OLS will provide poor out-of-sample performance. To take advantage of this wealth of data, other methods must be used.

This section focuses on one such approach, which uses the principal components of the data set to reduce the number of coefficients to be estimated. The use of principal components for prediction was discussed in Section 14.5; that treatment is extended here to time series data. The framework for doing so is the dynamic factor model (DFM), which models the comovements of a large number of time series as arising from a small number of unobserved variables, the so-called dynamic factors. One of the steps in estimating a DFM is estimation of these unobserved factors using principal components. As discussed at the end of this section, the DFM can be used for purposes other than forecasting.

The DFM is a widely used approach for forecasting with many time series predictors, but it is not the only approach. Another method is to estimate a VAR with many predictors but to use shrinkage methods, including Bayesian methods, to estimate those coefficients. For a graduate textbook discussion of Bayesian estimation of VARs, see Kilian and Lütkepohl (2017).

The Dynamic Factor Model

A central empirical regularity of developed economies is that there are broad common movements among macroeconomic variables: When there is strength in one part of the economy, there often is strength in other parts as well. At a horizon of several years, the common swings in many economic variables give rise to what are referred to as business cycles. Macroeconomic variables also move together at shorter horizons (months or quarters) and at longer horizons (decadal movements in long-term growth rates). Theories of macroeconomic fluctuations build on this empirical regularity of broadly observed comovements and attribute these comovements to a relatively small number of driving forces, such as productivity improvements, monetary policy, fiscal policy, and changes in demand or consumer preferences.

The **dynamic factor model** captures this notion that there are a small number (r) of common factors, which drive the comovements among a large number (N) of time series variables. The DFM treats these driving factors as unobserved. Treating the factors as unobserved admits that macroeconomists do not know all the sources of macroeconomic fluctuations and that even if they did, those sources would be difficult to measure directly (for example, technological progress is very difficult to measure). In a DFM, observed macroeconomic variables, such as GDP growth and the unemployment rate, are modeled as depending on these common unobserved factors and on other omitted drivers or measurement error.

Stated mathematically, the DFM has two parts. The first relates each of the N observable variables, X_{it}, to the r factors $F_{1t}, \ldots, F_{rt}$ plus an error term u_{it}:

$$X_{it} = \Lambda_{i0} + \Lambda_{i1}F_{1t} + \cdots + \Lambda_{ir}F_{rt} + u_{it}, i = 1, \ldots, N, \tag{17.32}$$

where $\Lambda_{i1}, \ldots, \Lambda_{ir}$ are unknown coefficients relating the r factors to the i^{th} observable variable and u_{it} is a mean 0 error term that represents omitted effects that are unique to X_{it} (that is, not common across variables) and measurement error.

The second part of the DFM specifies that the r factors follow a VAR. For notational convenience, we write the VAR here with a single lag [that is, as a VAR(1)]; however, more lags can be included:

$$\begin{aligned} F_{1t} &= \mathrm{A}_{11}F_{1t-1} + \mathrm{A}_{12}F_{2t-1} + \cdots + \mathrm{A}_{1r}F_{rt-1} + \eta_{1t} \\ &\vdots \\ F_{rt} &= \mathrm{A}_{r1}F_{1t-1} + \mathrm{A}_{r2}F_{2t-1} + \cdots + \mathrm{A}_{rr}F_{rt-1} + \eta_{rt}, \end{aligned} \tag{17.33}$$

where the A's are unknown VAR coefficients and the η's are mean 0 error terms. The factor VAR in Equation (17.33) is the extension to multiple variables (the r factors) of the two-variable VAR in Key Concept 17.1.

The error term u_{it} is assumed to be uncorrelated across series and to be uncorrelated with the factor VAR errors—that is, $E(u_{it}u_{jt+k}) = 0,\ i \neq j$, and $E(u_{it}\eta_{jt+k}) = 0$ for all k—so that all the common movements are associated with the common factors. Because there is no intercept in Equation (17.33), the factors have mean 0.

The **common component** of X_t is the part of X_{it} that is explained by the factors—that is, the predicted value of X_{it} given the factors, based on the population coefficients. In Equation (17.32), it is $\Lambda_{i1}F_{1t} + \cdots + \Lambda_{ir}F_{ir}$. The error term in Equation (17.32), u_{it}, is called the **idiosyncratic component** of X_{it} because it is the part of X_{it} not explained by the common factors. In general, the idiosyncratic component can be serially correlated, which affects how forecasts are made using the DFM.[3]

The DFM: Estimation and Forecasting

From the perspective of forecasting with many predictors, the DFM resolves the problem of having many predictors by replacing the many available time series with a small number of factors. If the factors were observed, the Λ coefficients in Equation (17.32) and the VAR coefficients in Equation (17.33) therefore could be estimated by OLS. The difficulty, however, is that the factors are not observed. The factors can, however, be estimated by the principal components of the N observed X's. These estimated factors can then be treated as data for the purpose of estimating the unknown DFM coefficients.

[3]Equations (17.32) and (17.33) are the so-called static form of the DFM, which is the version of the DFM most directly amenable to principal components estimation. Other forms of the DFM, and other ways to estimate the factors, are discussed in Stock and Watson (2016).

Estimation of the DFM and the factors using principal components. The method of principal components described in Section 14.5 extends directly to the time series setting. As discussed in Section 14.5, the X variables must first be standardized using their in-sample means and standard deviations; then the principal components are computed using the standardized X's. In Section 14.5, the first r principal components were denoted $PC_1, \ldots, PC_r$. In the context of the DFM, these principal components are the estimates of the common factors, and their value at date t is denoted $\hat{F}_{1t}, \ldots, \hat{F}_{rt}$, where the caret (^) indicates that the factor is estimated. If the factor model assumptions are, in fact, correct, then the principal components are consistent estimates of the factors in the sense that predictions made using the factors (were they observed) and using the principal components will be the same when both N and T are large.

Given the estimated factors $\hat{F}_{1t}, \ldots, \hat{F}_{rt}$, the Λ and A coefficients of the DFM in Equations (17.32) and (17.33) can be estimated by OLS, where the estimated factors are treated as data.

It is tempting to interpret the principal components themselves; for example, one might want to interpret the first principal component (the first estimated factor) as measuring overall economic activity, the second as measuring price inflation, and so forth. Unfortunately, such interpretations generally are not justified. The reason is that the factors are identified only up to linear combinations; without further assumptions, the factors themselves are not identified. Said differently, the common components of the series are identified in the dynamic factor model, but the factors themselves are not. For forecasting, this identification issue is irrelevant because the same forecasts will arise whether the factors or a linear combination of them is used (recall that, with OLS, the same prediction is made using, say, an intercept and the binary variable *male* as with an intercept and the binary variable *female*).

Determining the number of factors. In Chapter 14, the number of principal components was determined by leave-m-out cross validation. This method entails randomly assigning data to the m subsamples and then estimating the coefficients on the m subsamples that omit those observations. Unfortunately, leave-m-out cross validation has two problems in time series data. First, the time series observations are not independent, so the omitted data in the left-out subsample are not independent of the estimation sample. Second, if a subsample, even a contiguous subsample, is omitted, additional observations are lost because of the lag structure in the model.

For these reasons, determining the number of factors for DFMs tends to rely on scree plots and information criteria.

The scree plot with time series data is the same as that with cross-sectional data and is explained in Section 14.5.

Information criteria for determining the number of factors in a DFM have a similar structure to those used to determine the lag length for an autoregression [Equation (15.23)] or for a VAR [Equation (17.4)]. Specifically, the information criterion penalizes the sum of squared residuals for adding another factor. The information criterion approach to determining r was introduced by Bai and Ng (2002). A specific criterion they propose, which has been found to work well in simulations, is

$$IC(r) = \ln\left\{\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}[X_{it} - (\hat{\Lambda}_{i0} + \hat{\Lambda}_{i1}\hat{F}_{1t} + \cdots + \hat{\Lambda}_{ir}\hat{F}_{rt})]^2\right\} + r\left(\frac{N + T}{NT}\right)\ln[\min(N, T)] \quad (17.34)$$

where the $\hat{\Lambda}$'s are the OLS estimates of the Λ's, estimated using the first r principal components as regressors, and the final term is the penalty for using r principal components.

The Bai–Ng penalty in Equation (17.34) increases proportionately to the number of factors r, with a constant of proportionality that depends on the number of variables as well as the number of time series observations. When $N = T$, this penalty simplifies to 2 times the BIC penalty, $[\ln(T)]/T$.

Estimation of the number of factors using the information criterion in Equation (17.34) proceeds as for autoregressions and VARs: Among a set of candidate values of r, the estimated lag length is the value of r that minimizes IC(r).

Forecasting using the estimated factors. There are two approaches to forecasting using the estimated factors, which parallel the iterated and direct approaches to multi-period forecasting described in Section 17.2.

The starting point for both approaches is to extend Equation (17.32) to an autoregressive distributed lag model. Because u_{it} is, in general, serially correlated, past values of u_{it} are useful for forecasting u_{it} and thus X_{it}. Accordingly, the argument leading to Equation (16.21) applies here, so that the serial correlation in u_{it} implies that lagged values of X_{it} might be useful predictors as well. With these lagged terms added, Equation (17.32) becomes

$$X_{it} = \Lambda_{i0} + \Lambda_{i1}F_{1t} + \cdots + \Lambda_{ir}F_{rt} + \beta_1 X_{it-1} + \cdots + \beta_p X_{it-p} + u_{it}. \quad (17.35)$$

The right-hand side of Equation (17.35) depends on $F_{1t}, \ldots, F_{rt}$, which are unknown at date $t - 1$; thus current values of the factors (or their principal components estimates) cannot be used as predictors. The iterated and direct forecasting approaches take two different tacks to address this problem.

In the iterated approach, the contemporaneous values of the factors in Equation (17.35) are replaced by their forecasts from the estimated factor VAR. Thus the one-step ahead forecast for period $T + 1$, using data through period T, is

$$\hat{X}_{iT+1|T} = \hat{\Lambda}_{i0} + \hat{\Lambda}_{i1}\hat{F}_{1T+1|T} + \cdots + \hat{\Lambda}_{ir}\hat{F}_{rT+1|T} + \hat{\beta}_1 X_{iT} + \cdots + \hat{\beta}_p X_{iT-p+1}, \quad (17.36)$$

where the $\hat{\Lambda}$'s and $\hat{\beta}$'s are the estimates of the Λ's and β's in Equation (17.32) using $\hat{F}_{1t}, \ldots, \hat{F}_{rt}$ and lagged X's as regressors and where $\hat{F}_{1T+1|T}, \ldots, \hat{F}_{rT+1|T}$ are the one-step ahead forecasts of the factors computed using the factor VAR. Forecasts for horizons $h > 1$ are computed using the iterated VAR forecasts of the factors and of X_i.

The direct approach builds on Key Concept 17.3. Specifically, the h-step ahead direct forecasting regression using the estimated factors is

$$X_{it} = \delta_0 + \delta_1 \hat{F}_{1t-h} + \cdots + \delta_r \hat{F}_{rt-h} + \delta_{r+1} X_{it-h} + \cdots + \delta_{r+p} X_{it-h-p} + u_{it}, \tag{17.37}$$

where there are different regressions, and thus different δ coefficients, at each forecasting horizon h. For a given horizon, the coefficients of Equation (17.37) can be estimated by OLS, and the direct forecasts are then made using those estimated coefficients.

Typically, the coefficients are estimated using data through a specific date, and then the coefficients are frozen and used for real-time forecasting. This introduces a subtlety for forecasting with DFMs: The final observations on the factors, which are used to make real-time forecasts, might not have appeared in the estimation data set. As discussed in Appendix 14.5, because the coefficients are estimated using the in-sample principal components, the same weights and standardizing means and variances must be used to construct the principal components in the out-of-sample period as were used in the estimation sample.

Other uses of DFMs. DFMs can be used for purposes other than forecasting.

One such use is to construct economic indexes. If one has a large number of similar series, it can be useful to have a single summary index that captures the common comovements. In this case, a model with a single factor can be appropriate. The estimate of the single factor (the first principal component) then summarizes the comovements of all the variables. This approach is commonly used to compute a coincident economic index from multiple measures of economic activity.

Another use of DFMs is to estimate the *current* value of a variable. This problem arises because economic data are typically released with a lag. For example, one might be interested in the change of employment in the current month, but those data will not be released until next month. The task of "forecasting" current values of economic data is called **nowcasting**. The main technical challenge of nowcasting is that data are released over the course of any month, so that the nowcasting model must be able to incorporate incoming data as they arrive. The DFM is well suited to doing so, but it must be adapted to handle missing observations, and those methods are beyond the scope of this book. The Federal Reserve Bank of New York uses a DFM to produce nowcasts of GDP, which it updates weekly based on that week's data.[4]

Application to U.S. Macroeconomic Data

We illustrate the estimation and use of the dynamic factor model using a data set comprised of 131 quarterly macroeconomic time series for the United States, spanning 1960:Q1–2017:Q4. The series are summarized in Table 17.2, with additional information provided in Appendix 17.1. The variables in the data set include standard

[4]The New York Fed GDP nowcasts are posted at https://www.newyorkfed.org/research/policy/nowcast.

TABLE 17.2 The Quarterly Macroeconomic Data Set

Category	Number of Series Used for Factor Estimation
National Income and Product Accounts	13
Industrial Production	8
Employment and Unemployment	30
Orders, Inventories, and Sales	6
Housing Starts and Permits	6
Prices	22
Productivity and Labor Earnings	5
Interest Rates	10
Money and Credit	6
International	8
Asset Prices, Wealth, Household Balance Sheets	10
Other	2
Oil Market Variables	5
Total	**131**

measures of economic activity, wage and price inflation, interest rates, and data on large markets of macroeconomic importance including housing and oil markets. The data were transformed to eliminate stochastic trends, typically by transforming to growth rates (as for GDP) or first differences (interest rates). These transformed data were then standardized by subtracting their sample mean and dividing by their sample standard deviation prior to estimation.

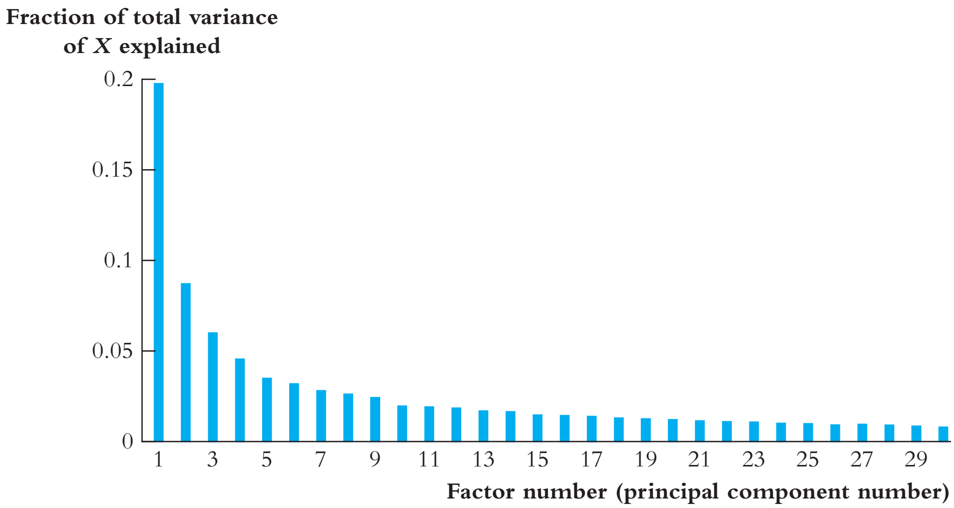
In some categories, series are available at multiple levels of aggregation. For example, GDP is the sum of consumption, investment, government spending, and imports; thus GDP is perfectly collinear with its components. Similarly, total employment is the sum of employment across the sectors of the economy. For the purpose of estimating the factors, the aggregate series (GDP, total employment) provide no additional information beyond their components, so the aggregate series were excluded from the data set. The final column of Table 17.2 lists the number of series used to compute the principal component factor estimates.

Figure 17.4 presents the scree plot of the first 30 principal components of the 131 series in the data set, over the full 1960–2017 period. Evidently, a large amount of the variance of these series is captured by the first few principal components. The first principal component explains 20% of the total variance of the series, the second principal component explains 9%, and the first four collectively explain 39%.

The scree plot provides some guidance about the number of factors to include. Clearly, the first and second factors are important, and there are also substantial drops in the marginal R^2 after the third and fourth factors. The decline does not seem to stabilize, however, until the tenth factor, so this visual analysis is inconclusive. The Bai–Ng information criterion [Equation (17.34)] is minimized using $r = 4$ factors.

FIGURE 17.4 Scree Plot of First 30 Factors for the Macro Data Set, 1960–2017

The first factor explains 20% of the total variance of the series, and the first four factors collectively explain 39% of the total variance of the series.



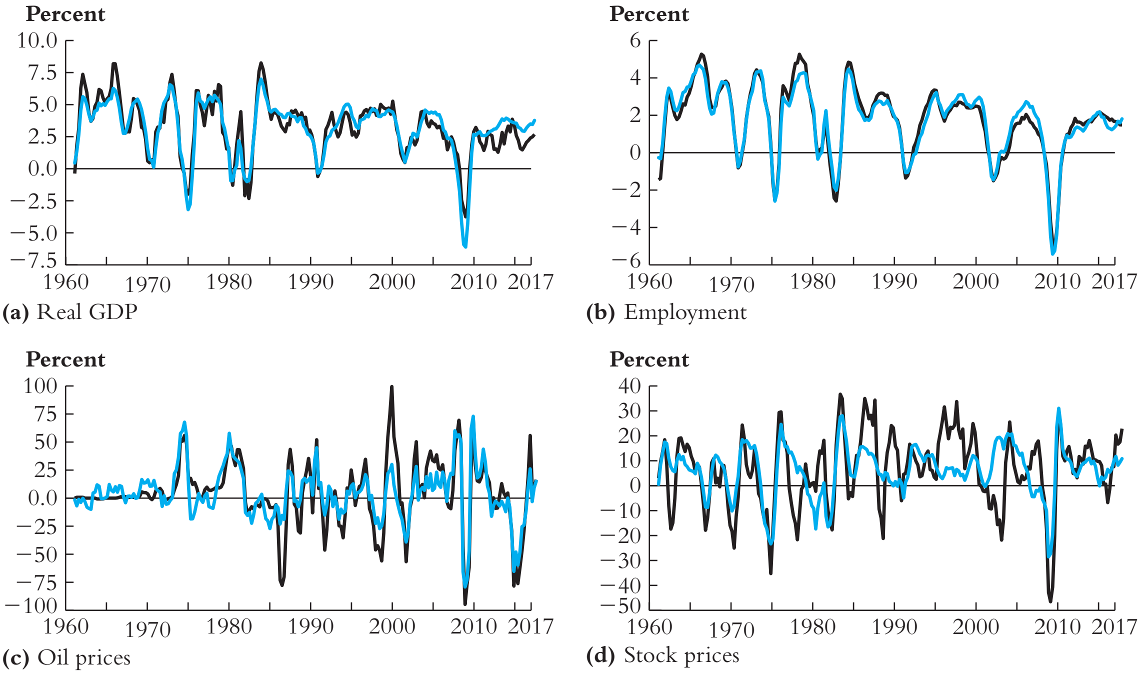
This estimate is within the plausible range from the inspection of the scree plot, so we adopt $r = 4$ for the rest of this example.

Figure 17.5 plots the four-quarter growth rate of GDP, employment, oil prices, and returns on the S&P 500 stock index (the four-quarter growth is the percentage growth of the series from quarter t to quarter $t + 4$, computed using the log approximation to percentage changes). The figure also plots the common component of each of the series, estimated using four factors. Of these series, GDP and employment are not in the data set used to estimate the factors because they are aggregates of other included series, while the oil prices and stock returns are among the 131 series used to estimate the factors.

The striking conclusion from Figure 17.5 is that the common component, computed using only the first 4 principal components of the 131 macro variables, captures a large amount of the variation in these series. Even a large fraction of the four-quarter returns on the S&P 500 are explained by these 4 factors. This does not imply that stock returns are predictable; rather, it implies that stock returns are heavily influenced by contemporaneous developments in aggregate economic activity.

We conclude by examining forecasts of GDP growth made using the four estimated factors and comparing those to the AR and ADL forecasts in Chapter 15. We consider direct forecasts of cumulative GDP growth at horizons $h = 1, 4$, and 8, where growth is measured at an annual rate. For example, at the four-quarter horizon, the dependent variable is $400\ln(GDP_t/GDP_{t-4})$, which equals the average of the quarterly growth in periods $t, t - 1, t - 2$, and $t - 3$ at an annual rate. The three forecasting models examined are direct forecasts of h-period growth corresponding to an AR(2), an ADL(2, 2) with the term spread, and a four-factor forecast that includes two lags of GDP growth.

FIGURE 17.5 Four-Quarter Growth Rates, Actual and Common Components, 1960–2017



The series (black) and estimated common components (blue) of GDP, employment, oil prices, and returns on the S&P 500 based on a four-factor DFM, estimated using the 131-series macroeconomic data set, 1960–2017.

Table 17.3 reports the performance of the forecasts as measured by the pseudo out-of-sample root mean square forecast error, $\widehat{RMSFE}_{POOS}$ [Equation (15.22)]. The first column lists the regressors in the direct forecasting regressions. Following Section 15.8, the in-sample period starts in 1981:Q1 and ends h periods prior to 2002:Q4; the pseudo out-of-sample period is 2002:Q4–2017:Q4.

Three aspects of these results are noteworthy. First, the $\widehat{RMSFE}_{POOS}$ decreases as the horizon lengthens. One reason for this improvement at longer horizons is that quarterly GDP has a large amount of transitory measurement error, which is smoothed over (averaged out) by considering growth rates over one or two years. This quarterly "noise" is evident in the time series plot of quarterly GDP growth in Figure 15.1b.

Second, at all horizons the forecasts that use the term spread do worse in the out-of-sample period than the direct AR(2) forecasts. This would appear to contradict the improvement in in-sample fit provided by the term spread: The F-statistic testing whether the coefficients on $TSpread_{t-1}$ and $TSpread_{t-2}$ are 0 in the $h = 1$ estimation sample (1981:Q1–2002:Q3) is statistically significant at the 1% level. Evidently, the coefficients on the lagged term spread estimated in the in-sample period do not capture the relation between the term spread and GDP in the pseudo out-of-sample period, an indication that this relation is nonstationary. In real-world terms, one important difference between the in- and out-of-sample periods is that, starting

TABLE 17.3 Comparison of Direct Forecasts of Cumulative GDP Growth at an Annual Rate: Lagged GDP, Term Spread, and Principal Components, 2002:Q4–2017:Q4

	$\widehat{RMSFE}_{POOS}$		
Predictors	$h = 1$	$h = 4$	$h = 8$
$GDPGR_{t-h}$, $GDPGR_{t-h-1}$	2.25	1.91	1.74
$GDPGR_{t-h}$, $GDPGR_{t-h-1}$, $TSpread_{t-h}$, $TSpread_{t-h-1}$	2.29	1.94	1.77
$GDPGR_{t-h}$, $GDPGR_{t-h-1}$, $\hat{F}_{1t-h}$, $\hat{F}_{2t-h}$, $\hat{F}_{3t-h}$, $\hat{F}_{4t-h}$	2.14	1.40	1.48

Entries are root mean square forecast errors, estimated by pseudo out-of-sample forecasts for the forecast period 2002:Q4–2017:Q4 [Equation (15.22)]. The forecasting models were estimated using data from 1981:Q1 through h periods before 2002:Q4, where h is the forecast horizon. The dependent variable is the h-quarter cumulative growth in GDP at an annual rate, using log points—that is, $(400/h)\ln(GDP_t/GDP_{t-h})$. The regressors are given in the first column, where $\hat{F}_{1t}$ denotes the first factor estimated by the first principal component in the estimation sample and so on. All regressions include an intercept.

in 2008, the Federal Reserve Board introduced new monetary policy tools to manage long-term as well as short-term rates, thereby changing the relation between the term spread and economic activity.

Third, the factor forecasts improve upon the AR and ADL forecasts at all horizons. Closer inspection of the forecasts reveals that this improvement is due to much better performance of the factor forecasts during the recession and early recovery following the financial crisis in the fall of 2009. During this recession, the strong negative comovements across many macro variables pointed toward a deep recession, a feature missed by the AR forecast. In contrast, during the relatively quiescent periods of 2005 and after 2013, the AR(2) direct forecast actually performs slightly better than the factor forecast.

Nobel Laureates in Time Series Econometrics

In 2003, Robert Engle and Clive Granger won the Nobel Prize in Economics for fundamental theoretical research in time series econometrics. Engle's work was motivated by the volatility clustering evident in plots like Figure 17.2. Engle wondered whether series like these could be stationary and whether econometric models could be developed to explain and predict their time-varying volatility. Engle's answer was to develop the autoregressive conditional heteroskedasticity (ARCH) model, described in Section 17.5. The ARCH model and its extensions proved especially useful for modeling the volatility of asset returns, and the resulting volatility forecasts are 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.

John McCombe/AP Images

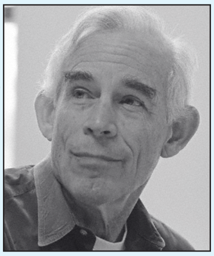
Clive W. J. Granger

New York University/AFP/Newscom

Robert F. Engle

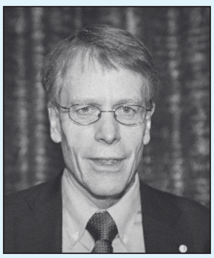
Granger's work focused on how to handle stochastic trends in economic time series data. From his earlier work, he knew that two unrelated series with stochastic trends could, by the usual statistical measures of t-statistics and regression R^2's, falsely appear to be meaningfully related; this is the "spurious regression" problem exemplified by the regressions in Equations (14.28) and (14.29). But are all regressions involving stochastic trending variables spurious? Granger discovered that when variables shared common trends—in his terminology, were "co-integrated"—meaningful relationships could be uncovered by regression analysis using a vector error correction model. The methods of cointegration analysis are now a staple in modern macroeconometrics.

In 2011, Thomas Sargent and Christopher Sims won the Nobel Prize for their empirical research on cause and effect in the macroeconomy. Sargent was recognized for developing models that featured the important role that expectations about the future play in disentangling cause and effect. Sims was recognized for developing structural VAR (SVAR) models. Sims's key insight concerned the forecast errors in a VAR model—the u_t errors in Equations (17.1) and (17.2). These errors, he realized, arose because of unforeseen "shocks" that buffeted the macroeconomy, and in many cases, these shocks had well-defined sources like the Organization of Petroleum Exporting Countries (oil price shocks), the Fed (interest rate shocks), or Congress (tax shocks). By disentangling the various sources of shocks that comprise the VAR errors, Sims was able to estimate the dynamic causal effect of these shocks on the variables appearing in the VAR. This disentangling of shocks is never without controversy, but SVARs are now a standard tool for estimating dynamic causal effects in macroeconomics.



Julio Cortez/AP Images

Christopher A. Sims



Karl Schoendorfer/Shutterstock

Lars Peter Hansen

In 2013, Eugene Fama, Lars Peter Hansen, and Robert Shiller won the Nobel Prize for their empirical analysis of asset prices. The work in the box "Can You Beat the Market?" in Chapter 15 and the box "NEWS FLASH: Commodity Traders Send Shivers Through Disney World" in Chapter 16 was motivated in part by the "efficient markets" (unpredictability) work of Fama and the "irrational exuberance" (unexplained volatility) work of Shiller. Hansen was honored for developing generalized method of moments (GMM) methods to investigate whether asset returns are consistent with expected utility theory. Microeconomics says that investors should equate the marginal cost of an investment (today's foregone utility from investing rather than consuming) with its marginal benefit (tomorrow's boost in utility from consumption financed by the investment's return). But a simple test of this proposition is complicated because marginal utility is difficult to measure, asset returns are uncertain, and the argument should hold across all asset returns. Hansen developed GMM methods to test asset-pricing models. As it turned out, Hansen's GMM methods had applications well beyond finance and are now widely used in econometrics. Section 19.7 introduces GMM.

For more information on these and other Nobel laureates in economics, visit the Nobel Foundation website, **http://www.nobel.se/economics**.

17.7 Conclusion

This part of the text 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 Diebold (2017) and Enders (2009). For an advanced treatment of econometrics with time series data, see Hamilton (1994) and Hayashi (2000). For an advanced treatment of vector autoregressions, see Kilian and Lütkepohl (2017), and for more on dynamic factor models, see Stock and Watson (2016).

Summary

1. Vector autoregressions model k time series variables, with each depending 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 multi-period 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 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—in which the variance of a series is high in some periods and low in others—is common in economic time series, especially financial time series. Realized volatility is an estimate of time-varying volatility using a rolling root mean square estimator.
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. Realized volatility and ARCH/GARCH models produce forecast intervals with widths that depend on the volatility of the most recent regression residuals.
6. The comovements of a large number of time series sometimes can be summarized by the first few principal components, which in turn can be used for forecasting. The framework for doing so is the dynamic factor model, which posits that a small number of unobserved factors drive the comovements of a large number of macroeconomic variables.

Key Terms

vector autoregression (VAR) (650)
iterated multi-period AR forecast (656)
iterated multi-period VAR forecast (656)
direct multi-period forecast (658)
second difference (659)
integrated of order zero [$I(0)$], one [$I(1)$], or two [$I(2)$] (659)
order of integration (659)
integrated of order d [$I(d)$] (660)
common trend (663)
cointegration (663)
cointegrating coefficient (664)
error correction term (663)
vector error correction model (VECM) (663)
EG-ADF test (665)
dynamic OLS (DOLS) estimator (665)
volatility clustering (667)
realized volatility (668)
autoregressive conditional heteroskedasticity (ARCH) (669)
generalized ARCH (GARCH) (669)
dynamic factor model (672)
common component (673)
idiosyncratic component (673)
nowcasting (676)

Review the Concepts

17.1 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 2017. Should he estimate a VAR for these variables and use this for forecasting? Why or why not? Can you suggest an alternative approach?

17.2 Suppose that Y_t follows a stationary AR(1) model with $\beta_0 = 0$ and $\beta_1 = 0.5$. If $Y_t = 10$, what is your forecast of Y_{t+2} (that is, what is $Y_{t+2|t}$)? What is $Y_{t+h|t}$ for $h = 20$? Does this forecast for $h = 20$ seem reasonable to you?

17.3 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.

17.4 What is volatility clustering? Explain two models that are used to describe data processes with volatility clustering.

17.5 What is a unit root? How does a researcher test for the presence of a unit root in the data?

Exercises

17.1 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 $X_t = [\mu_Y/(1 - \delta)] + [(Y_t - \mu_Y)/(1 - \beta_1\delta)]$.

17.2 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?

17.3 Suppose that $E(u_t | u_{t-1}, u_{t-2}, \ldots) = 0$ and u_t follows the ARCH process, $\sigma_t^2 = 1.0 + 0.5\, u_{t-1}^2$.

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, $E(u_t^2) = E[Eu_t^2 | u_{t-1})]$.)

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 $\Pr(-3 \le u_t \le 3)$?

17.4 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 $h > p$.

17.5 Verify Equation (17.20). [*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 $\sum_{t=1}^{T} Y_{t-1}\Delta Y_t$.]

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

$$Y_t = 2.0 + 1.5X_{t+1} + 0.9X_t - 0.3X_{t-1} + u_t.$$

a. Rearrange the regression so that it has the form shown in Equation (17.25). What are the values of θ, δ_{-1}, δ_0, and δ_1?

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

ii. Suppose that X_t is $I(1)$ and Y_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?

17.7 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 (17.21). Analyze the denominator using the law of large numbers.]

17.8 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 multi-period forecasts differ from direct multi-period forecasts? Explain.

17.9 **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.

17.10 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 0 serially uncorrelated random variables with $E(v_{1t} v_{2j}) = 0$ for all t and j. Derive the vector error correction model [Equations (17.22) and (17.23)] for X and Y.

Empirical Exercises

E17.1 This exercise is an extension of Empirical Exercise 14.1. On the text website, **http://www.pearsonglobaleditions.com**, you will find the data file **USMacro_Quarterly**, which contains quarterly data on several macroeconomic series for the United States; the data are described in the file **USMacro_Description**. Compute inflation, *Infl*, using the price index for personal consumption expenditures. For all regressions, use the sample period 1963:Q1–2017:Q4 (where data before 1963 may be used as initial values for lags in regressions).

a. Using the data on inflation through 2017:Q4 and an estimated AR(2) model:

i. Forecast $\Delta Infl_{2018:Q1}$, the change in inflation from 2017:Q4 to 2018:Q1.

ii. Forecast $\Delta Infl_{2018:Q2}$, the change in inflation from 2018:Q1 to 2018:Q2. (Use an iterated forecast.)

iii. Forecast $Infl_{2018:Q2} - Infl_{2017:Q4}$, the change in inflation from 2017:Q4 to 2018:Q2.

iv. Forecast $Infl_{2018:Q2}$, the rate of inflation in 2018:Q2.

b. Repeat (a) using the direct forecasting method.

E17.2 On the text website, **http://www.pearsonglobaleditions.com,** you will find the data file **USMacro_Quarterly**, which contains quarterly data on real GDP, measured in 2009 dollars. Compute $GDPGR_t = 400 \times [\ln(GDP_t) - \ln(GDP_{t-1})]$, the GDP growth rate.

a. Using data on $GDPGR_t$ from 1960:Q1 to 2017:Q4, estimate an AR(2) model with GARCH(1, 1) errors.

b. Plot the residuals from the AR(2) model along with $\pm\hat{\sigma}_t$ bands as in Figure 17.3.

c. Some macroeconomists have claimed that there was a sharp drop in the variability of the growth rate of GDP around 1983, which they call the Great Moderation. Is this Great Moderation evident in your plot for (b)? Explain.

APPENDIX

17.1 The Quarterly U.S. Macro Data Set

The variables in the quarterly U.S. data set were obtained from the FRED online database of macroeconomic time series maintained by the Federal Reserve Bank of St. Louis. The categories of variables are listed in Table 17.2. The National Income and Product Account variables included in the data set for estimating the factors are three measures of personal consumption expenditures (durable goods, nondurable goods, and services); four measures of private investment (nonresidential structures, nonresidential intellectual property, nonresidential fixed equipment, and residential structures), federal government expenditures, federal government receipts, state and local government consumption, exports, and imports (all real). Stochastic trends were eliminated by (in most cases) computing quarterly growth rates or first differences. For details and for the full list of series, see the online documentation supporting this text.