CHAPTER TWENTY-SEVEN

Forecasting Economic Time Series

James H. Stock*

1 INTRODUCTION

The construction and interpretation of economic forecasts is one of the most publicly visible activities of professional economists. Over the past two decades, increased computer power has made increasingly sophisticated forecasting methods routinely available and the role of economic forecasting has expanded. Economic forecasts now enter into many aspects of economic life, including business planning, state and local budgeting, financial management, financial engineering, and monetary and fiscal policy. Yet, with this widening scope comes greater opportunities for the production of poor forecasts and the misinterpretation of good forecasts. Responsible production and interpretation of economic forecasts requires a clear understanding of the associated econometric tools, their limits, and an awareness of common pitfalls in their application.

   This chapter provides an introduction to the main methods used for forecasting economic time series. The field of economic forecasting is large, and, because of space limitations, this chapter covers only the most salient topics. The focus here will be on point forecasts, that is, forecasts of future values of the time series. It is assumed that the historical series is relatively "clean," in the sense of having no omitted observations, being observed at a consistent sampling frequency (e.g. monthly), and either having no seasonal component or having been seasonally adjusted. It is assumed that the forecaster has quadratic (i.e. mean squared error) loss. Finally, it is assumed that the time series is sufficiently long, relative to the forecast horizon, that the history of the time series will be informative for making the forecast and for estimating parametric models.

   This chapter has four substantive sections. Section 2 provides a theoretical framework for considering some of the tradeoffs in the construction of economic forecasts and for the comparison of forecasting methods. Section 3 provides a

glimpse at some of the relevant empirical features of macroeconomic time series data. Section 4 discusses univariate forecasts, that is, forecasts of a series made using only past values of that series. Section 5 provides an overview of multivariate forecasting, in which forecasts are made using historical information on multiple time series.

There are many interesting and important aspects of economic forecasting that are not covered in this chapter. In some applications, it is of interest to estimate the entire distribution of future values of the variable of interest, conditional on current information, or certain functions of that conditional distribution. An example that arises in macroeconomics is predicting the probability of a recession, an event often modeled as two consecutive declines in real gross domestic product. Other functions of conditional distributions arise in finance; for examples, see Diebold, Gunther, and Tay (1998).

In some cases, time varying conditional densities might be adequately summarized by time varying conditional first and second moments, that is, by modeling conditional heteroskedasticity. For example, conditional estimates of future second moments of the returns on an asset can be used to price options written on that asset. Although there are various frameworks for estimating conditional heteroskedasticity, the premier tool for modeling conditional heteroskedasticity is Engle's (1982) so-called autoregressive conditional heteroskedasticity (ARCH) framework and variants, as discussed in Bollerslev, Engle, and Nelson (1994).

Another topic not explored in this chapter is nonquadratic loss. Quadratic loss is a natural starting point for many forecasting problems, both because of its tractability and because, in many applications, it is plausible that loss is symmetric and that the marginal cost of a forecast error increases linearly with its magnitude. However, in some circumstances other loss functions are appropriate. For example, loss might be asymmetric (would you rather be held responsible for a surprise government surplus or deficit?); see Granger and Newbold (1986, ch. 4.2) and West, Edison, and Cho (1993) for examples. Handling nonquadratic loss can be computationally challenging. The classic paper in this literature is Granger (1969), and a recent contribution is Christoffersen and Diebold (1997).

Another important set of problems encountered in practice but not addressed here involve data irregularities, such as missing or irregularly spaced observations. Methods for handling these irregularities tend to be model-dependent. Within univariate linear models and low-dimensional multivariate linear models, these are typically well handled using state space representations and the Kalman filter, as is detailed by Harvey (1989). A somewhat different set of issues arise with series that have large seasonal components. Issues of seasonal adjustment and handling seasonal data are discussed in Chapter 31 in this volume by Ghysels, Osborn, and Rodrigues.

Different issues also arise if the forecast horizon is long relative to the sample size (say, at least one-fifth the sample size) and the data exhibit strong serial correlation. Then the long-run forecast is dominated by estimates of the long-run correlation structure. Inference about the long-run correlation structure is typically nonstandard and, in some formulations, is related to the presence of large, possibly unit autoregressive roots and (in the multivariate setting) to possible cointegration

among the series. Unit roots and cointegration are respectively discussed in this volume in Chapter 29 by Bierens and Chapter 30 by Dolado, Gonzalo, and Marmol. The construction of point forecasts and forecast intervals at long horizons entails considerable difficulties because of the sensitivity to the long-run dependence parameters, and methods for handling this are examined in Stock (1996).

A final area not addressed here is the combination of competing forecasts. When a variable is forecasted by two different methods that draw on different information sets and neither model is true, typically a combination of the two forecasts is theoretically preferred to either individual forecast (Bates and Granger, 1969). For an introduction to this literature, see Granger (1989), Diebold and Lopez (1995), and Chan, Stock, and Watson (1998).

This chapter makes use of concepts and methods associated with unit autoregressive roots, cointegration, vector autoregressions (VARs), and structural breaks. These are all topics of separate chapters in this volume, and the reader is referred to those chapters for background details.

2   Economic Forecasting: A Theoretical Framework

2.1   Optimal forecasts, feasible forecasts, and forecast errors

Let y_t denote the scalar time series variable that the forecaster wishes to forecast, let h denote the horizon of the forecast, and let F_t denote the set of data used at time t for making the forecast (F_t is sometimes referred to as the information set available to the forecaster). If the forecaster has squared error loss, the point forecast $\hat{y}_{t+h|t}$ is the function of F_t that minimizes the expected squared forecast error, that is, $E[(y_{t+h} - \hat{y}_{t+h|t})^2 | F_t]$. This expected loss is minimized when the forecast is the conditional expectation, $E(y_{t+h}|F_t)$. In general, this conditional expectation might be a time varying function of F_t. However, in this chapter we will assume that the data are drawn from a stationary distribution, that is, the distribution of $(y_s, \ldots, y_{s+T})$ does not depend on s (although some mention of structural breaks will be made later); then $E(y_{t+h}|F_t)$ is a time invariant function of F_t.

In practice, $E(y_{t+h}|F_t)$ is unknown and is in general nonlinear. Forecasts are constructed by approximating this unknown conditional expectation by a parametric function. This parametric function, or model, is denoted $\mu_h(F_t, \theta)$, where θ is a parameter vector which is assumed to lie in the parameter space Θ. The "best" value of this parameter, θ_0, is the value that minimizes the mean squared approximation error, $E[\mu_h(F_t, \theta) - E(y_{t+h}|F_t)]^2$.

Because θ_0 is unknown, it is typically estimated from historical data, and the estimate is denoted $\hat{\theta}$. To be concrete, suppose that F_t consists of observations on X_s, $1 \leq s \leq T$, where X_s is a vector time series (which typically includes y_s). Further suppose that only the first p lags of X_t are included in the forecast. Then θ could be estimated by least squares, that is, by solving,

$$\min_{\theta \in \Theta} \sum_{t=p+1}^{T-h} [y_{t+h} - \mu_h(F_t, \theta)]^2. \tag{27.1}$$

There are alternative methods for the estimation of θ. The minimization problem (27.1) uses an *h*-step ahead (nonlinear) least squares regression. Often an available alternative is to estimate a one-step ahead model ($h = 1$) by nonlinear least squares or maximum likelihood, and to iterate that model forward. The formulation (27.1) has the advantage of computational simplicity, especially for nonlinear models. Depending on the true model and the approximate model, approximation bias can be reduced by estimating the *h*-step ahead model (27.1). On the other hand, if the estimated model is correct, then iterating one-step ahead forecasts will be more efficient in the statistical sense. In general, the decision of whether to estimate parameters by *h*-step ahead or one-step ahead methods depends on the model being estimated and the type of misspecification that might be present. See Clements and Hendry (1996) for references to this literature and for simulation results comparing the two approaches.

It is useful to consider a decomposition of the forecast error, based on the various sources of that error. Let $\hat{e}_{t+h,t}$ denote the forecast error from the *h*-step ahead forecast of y_{t+h} using $\hat{y}_{t+h|t}$. Then,

$$\hat{e}_{t+h,t} = y_{t+h} - \hat{y}_{t+h|t}$$
$$= [y_{t+h} - E(y_{t+h} | F_t)] + [E(y_{t+h} | F_t) - \mu_h(F_t, \theta_0)] + [\mu_h(F_t, \theta_0) - \mu_h(F_t, \hat{\theta})].$$
(27.2)

The first term in brackets is the deviation of y_{t+h} from its conditional expectation, a source of forecast error that cannot be eliminated. The second term in brackets is the contribution of model misspecification, and is the error arising from using the best parameter value for the approximate conditional expectations function. The final term arises because this best parameter value is unknown, and instead θ is estimated from the data.

The decomposition (27.2) illustrates two facts. First, all forecasts, no matter how good, will have forecast error because of future, unknowable random events. Second, the quality of a forecasting method is therefore determined by its model approximation error and by its estimation error. These two sources of error generally entail a tradeoff. Using a flexible model with many parameters for μ_h can reduce model approximation error, but because there are many parameters estimation error increases.

2.2 Model selection using information criteria

Because the object of point forecasting is to minimize expected loss out-of-sample, it is not desirable to minimize approximation error (bias) when this entails adding considerable parameter estimation uncertainty. Thus, for example, model selection based on minimizing the sum of squared residuals, or maximizing the R^2, can lead to small bias and good in-sample fit, but very poor out-of-sample forecast performance.

A formal way to make this tradeoff between approximation error and estimation error is to use information criteria to select among a few competing models. When $h = 1$, information criteria (IC) have the form,

$$IC(p) = \ln \hat{\sigma}^2(p) + pg(T) \tag{27.3}$$

where p is the dimension of θ, T is the sample size used for estimation, $g(T)$ is a function of T with $g(T) > 0$ and $Tg(T) \to \infty$ and $g(T) \to 0$ as $T \to \infty$, and $\hat{\sigma}^2(p) = \text{SSR}/T$, where SSR is the sum of squared residuals from the (in-sample) estimation. Comparing two models using the information criterion (27.3) is the same as comparing two models by their sum of squared residuals, except that the model with more parameters receives a penalty. Under suitable conditions on this penalty and on the class of models being considered, it can be shown that a model selected by the information criterion is the best in the sense of the trade-off between approximation error and sampling uncertainty about θ. A precise statement of such conditions in AR models, when only the maximum order is known, can be found in Geweke and Meese (1981), and extensions to infinite order autoregressive models are discussed in Brockwell and Davis (1987) and, in the context of unit root tests, Ng and Perron (1995). The two most common information criteria are the Akaike information criterion (AIC), for which $g(T) = 2/T$, and Schwarz's (1978) Bayes information criterion (BIC), for which $g(T) = \ln T/T$.

2.3 Prediction intervals

In some cases, the object of forecasting is not to produce a point forecast but rather to produce a range within which y_{t+h} has a prespecified probability of falling. Even if within the context of point forecasting, it is useful to provide users of forecasts with a measure of the uncertainty of the forecast. Both ends can be accomplished by reporting prediction intervals.

In general, the form of the prediction interval depends on the underlying distribution of the data. The simplest prediction interval is obtained by assuming that the data are conditionally homoskedastic and normal. Under these assumptions and regularity conditions, a prediction interval with asymptotic 67 percent coverage is given by $\hat{y}_{t+h|t} \pm \tilde{\sigma}_h$, where $\tilde{\sigma}_h = \text{SSR}_h/(T - p)$, where SSR_h is the sum of squared residuals from the h-step ahead regression (27.1) and $T - p$ are the degrees of freedom of that regression.

If the series is conditionally normal but is conditionally heteroskedastic, this simple prediction error formula must be modified and the conditional variance can be computed using, for example, an ARCH model. If the series is conditionally nonnormally distributed, other methods, such as the bootstrap, can be used to construct asymptotically valid prediction intervals.

2.4 Forecast comparison and evaluation

The most reliable way to evaluate a forecast or to compare forecasting methods is by examining out of sample performance. To evaluate the forecasting performance of a single model or expert, one looks for signs of internal consistency. If the

forecasts were made under squared error loss, the forecast errors should have mean zero and should be uncorrelated with any variable used to produce the forecast. For example, $\hat{e}_{t+h,t}$ should be uncorrelated with $\hat{e}_{t,t-h}$, although $\hat{e}_{t+h,t}$ will in general have an MA($h - 1$) correlation structure. Failure of out-of-sample forecasts to have mean zero and to be uncorrelated with F_t indicates a structural break, a deficiency of the forecasting model, or both.

Additional insights are obtained by comparing the out-of-sample forecasts of competing models or experts. Under mean squared error loss, the relative performance of two time series of point forecasts of the same variable can be compared by computing their mean squared forecast errors (MSFE). Of course, in a finite sample, a smaller MSFE might simply be an artifact of sampling error, so formal tests of whether the MSFEs are statistically significantly different are in order when comparing two forecasts. Such tests have been developed by Diebold and Mariano (1995) and further refined by West (1996), who built on earlier work by Nelson (1972), Fair (1980), and others.

Out of sample performance can be measured either by using true out-of-sample forecasts, or by a simulated out-of-sample forecasting exercise. While both approaches have similar objectives, the practical issues and interpretation of results is quite different. Because real time published forecasts usually involve expert opinion, a comparison of true out-of-sample forecasts typically entails an evaluation of both models and the expertise of those who use the models. Good examples of comparisons of real time forecasts, and of the lessons that can be drawn from such comparisons, are McNees (1990) and Zarnowitz and Braun (1993).

Simulated real time forecasting can be done in the course of model development and provides a useful check on the in-sample comparison measures discussed above. The essence of a simulated real time forecasting experiment is that all forecasts $\hat{y}_{t+h|t}$, $t = T_0, \ldots, T_1$, are functions only of data up through date t, so that all parameter estimation, model selection, etc., is done only using data through date t. This is often referred to as a recursive methodology (for linear models, the simulated out-of-sample forecasts can be computed using a recursion). In general this entails many re-estimations of the model, which for nonlinear models can be computationally demanding. For an example of simulated out-of-sample forecast comparisons, see Stock and Watson (1999a).

3   SALIENT FEATURES OF US MACROECONOMIC TIME SERIES DATA

The methods discussed in this chapter will be illustrated by application to five monthly economic time series for the US macroeconomy: inflation, as measured by the annual percentage change in the consumer price index (CPI); output growth, as measured by the growth rate of the index of industrial production; the unemployment rate; a short-term interest rate, as measured by the rate on 90-day US Treasury bills; and total real manufacturing and trade inventories, in logarithms.[1] Time series plots of these five series are presented as the heavy solid lines in Figures 27.1–27.5.
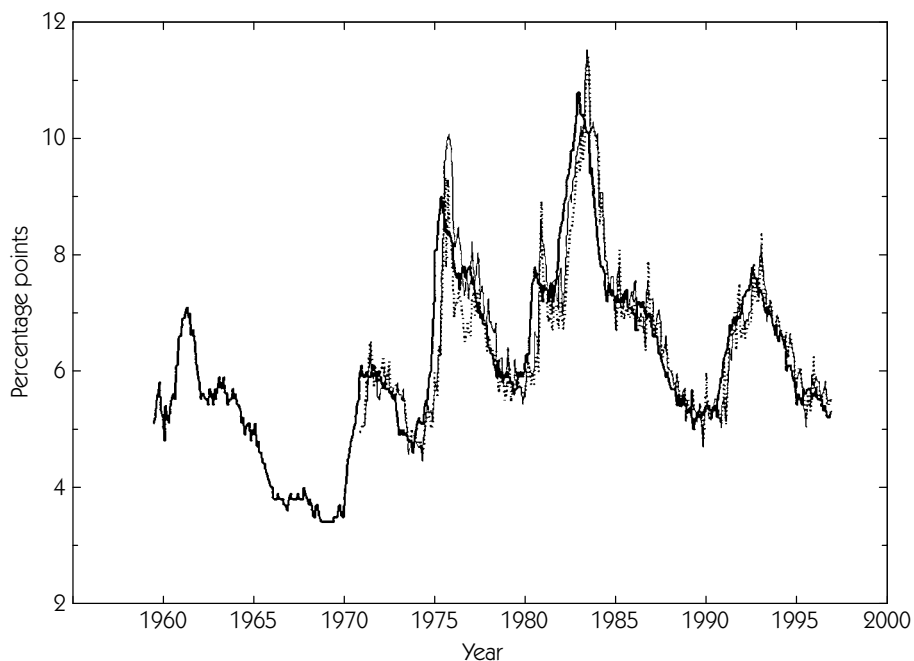
Figure 27.1 US unemployment rate (heavy solid line), recursive AR(BIC)/unit root pretest forecast (light solid line), and neural network forecast (dotted line)
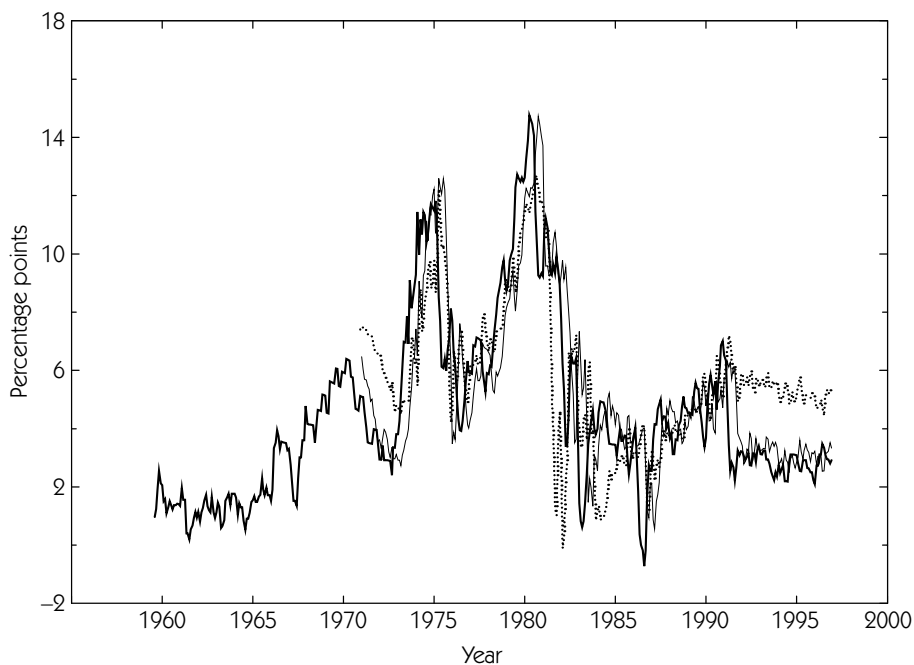

Figure 27.2 Six-month US CPI inflation at an annual rate (heavy solid line), recursive AR(BIC)/unit root pretest forecast (light solid line), and neural network forecast (dotted line)

Figure 27.3 90-day Treasury bill at an annual rate (heavy solid line), recursive AR(BIC)/ unit root pretest forecast (light solid line), and neural network forecast (dotted line)

Figure 27.4 Six-month growth of US industrial production at an annual rate (heavy solid line), recursive AR(BIC)/unit root pretest forecast (light solid line), and neural network forecast (dotted line)
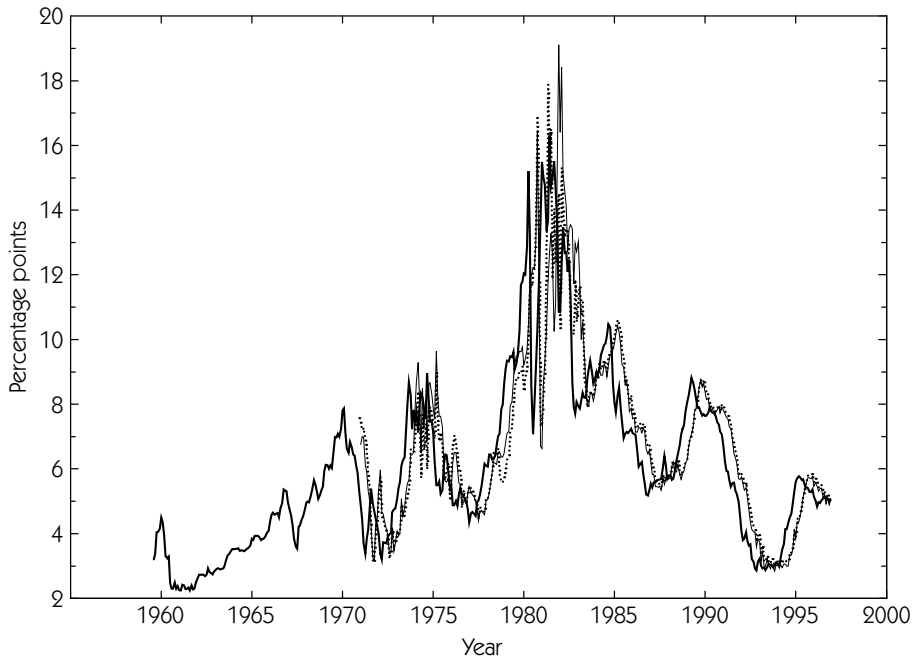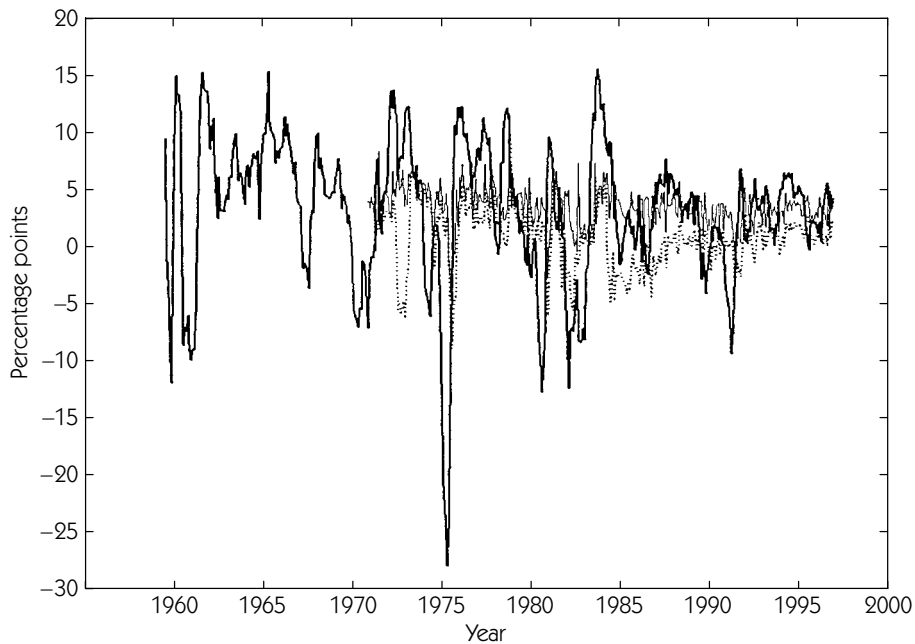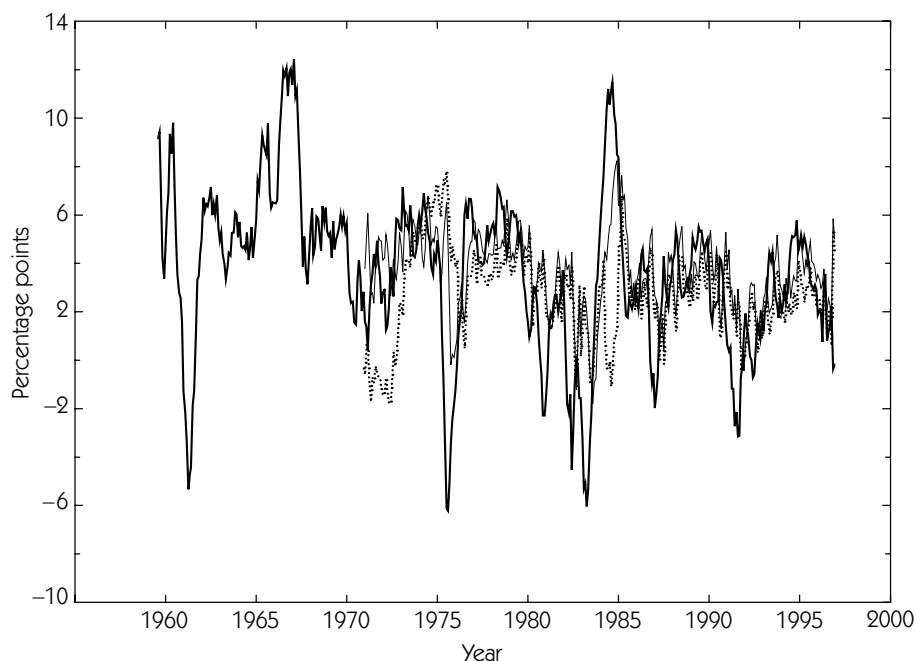
Figure 27.5 Six-month growth of total real US manufacturing and trade inventories at an annual rate (heavy solid line), recursive AR(BIC)/unit root pretest forecast (light solid line), and neural network forecast (dotted line)

In addition to being of interest in their own right, these series reflect some of the main statistical features present in many macroeconomic time series from developed economies. The 90-day Treasury bill rate, unemployment, inflation, and inventories all exhibit high persistence in the form of smooth long-run trends. These trends are clearly nonlinear, however, and follow no evident deterministic form, rather, the long-run component of these series can be thought of as a highly persistent stochastic trend. There has been much debate over whether this persistence is well modeled as arising from an autoregressive unit root in these series, and the issue of whether to impose a unit root (to first difference these data) is an important forecasting decision discussed below.

Two other features are evident in these series. All five series exhibit co-movements, especially over the two to four year horizons. The twin recessions of the early 1980s, the long expansions of the mid-1980s and the 1990s, and the recession in 1990 are reflected in each series (although the IP (industrial production) growth rate series might require some smoothing to see this). Such movements over the business cycle are typical for macroeconomic time series data; for further discussion of business cycle properties of economic time series data, see Stock and Watson (1999b). Finally, to varying degrees the series contain high frequency noise. This is most evident in inflation and IP growth. This high frequency noise arises from short-term, essentially random fluctuations in economic activity and from measurement error.

4   UNIVARIATE FORECASTS

Univariate forecasts are made solely using past observations on the series being forecast. Even if economic theory suggests additional variables that should be useful in forecasting a particular variable, univariate forecasts provide a simple and often reliable benchmark against which to assess the performance of those multivariate methods. In this section, some linear and nonlinear univariate forecasting methods are briefly presented. The performance of these methods is then illustrated for the macroeconomic time series in Figures 27.1–27.5.

4.1   Linear models

One of the simplest forecasting methods is the *exponential smoothing* or exponentially weighted moving average (EWMA) method. The EWMA forecast is,

$$\hat{y}_{t+h|t} = \alpha \hat{y}_{t+h-1|t-1} + (1 - \alpha)y_t, \tag{27.4}$$

where α is a parameter chosen by the forecaster or estimated by nonlinear least squares from historical data.

Autoregressive moving average (ARMA) models are a mainstay of univariate forecasting. The ARMA(p, q) model is,

$$a(L)y_t = \mu_t + b(L)\varepsilon_t, \tag{27.5}$$

where ε_t is a serially uncorrelated disturbance and $a(L)$ and $b(L)$ are lag polynomials of orders p and q, respectively. For y_t to be stationary, the roots of $a(L)$ lie outside the unit circle, and for $b(L)$ to be invertible, the roots of $b(L)$ also lie outside the unit circle. The term μ_t summarizes the deterministic component of the series. For example, if μ_t is a constant, the series is stationary around a constant mean. If $\mu_t = \mu_0 + \mu_1 t$, the series is stationary around a linear time trend. If $q > 0$, estimation of the unkown parameters of $a(L)$ and $b(L)$ entails nonlinear maximization. Asymptotic Gaussian maximum likelihood estimates of these parameters are a staple of time series forecasting computer packages. Multistep forecasts are computed by iterating forward the one-step forecasts. A deficiency of ARMA models is estimator bias introduced when the MA roots are large, the so-called unit MA root pileup problem (see Davis and Dunsmuir, 1996; and, for a general discussion and references, Stock, 1994).

An important special case of ARMA models are pure *autoregressive* models with lag order p (AR(p)). Meese and Geweke (1984) performed a large simulated out of sample forecasting comparison that examined a variety of linear forecasts, and found that long autoregressions and autoregressions with lags selected by information criteria performed well, and on average outperformed forecasts from ARMA models. The parameters can be estimated by ordinary least squares (OLS) and the order of the autoregression can be consistently estimated by, for example, the BIC.

Harvey (1989) has proposed a different framework for univariate forecasting, based on a decomposition of a series into various components: trend, cycle, seasonal, and irregular. Conceptually, this framework draws on an old concept in economic time series analysis in which the series is thought of as having different properties at different horizons, so that for example one might talk about the cyclical properties of a time series separately from its trend properties; he therefore calls these *structural time series models*. Harvey models these components as statistically uncorrelated at all leads and lags, and he parameterizes the components to reflect their role, for example, the trend can be modeled as a random walk with drift or a doubly integrated random walk, possibly with drift. Estimation is by asymptotic Gaussian maximum likelihood. The resulting forecasts are linear in historical data (although nonlinear in the parameters of the model) so these too are linear forecasts. Harvey (1989) argues that this formulation produces forecasts that avoid some of the undesirable properties of ARMA models. As with ARMA models, user judgment is required to select the models. One interesting application of these models is for trend estimation, see for example Stock and Watson (1998).

4.2   Nonlinear models

Outside of the normal distribution, conditional expectations are typically nonlinear, and in general one would imagine that these infeasible optimal forecasts would be nonlinear functions of past data. The main difficulty that arises with nonlinear forecasts is choosing a feasible forecasting method that performs well with the fairly short historical time series available for macroeconomic forecasting. With many parameters, approximation error in (27.2) is reduced, but estimation error can be increased. Many nonlinear forecasting methods also pose technical problems, such as having objective functions with many local minima, having parameters that are not globally identified, and difficulties with generating internally consistent h-step ahead forecasts from one-step ahead models.

Recognition of these issues has led to the development of a vast array of methods for nonlinear forecasting, and a comprehensive survey of these methods is beyond the limited scope of this chapter. Rather, here I provide a brief introduction to only two particular nonlinear models, smooth transition autoregressions (STAR) and artificial neural networks (NN). These models are interesting methodologically because they represent, respectively, parametric and nonparametric approaches to nonlinear forecasting, and they are interesting from a practical point of view because they have been fairly widely applied to economic data.

A third class of models that has received considerable attention in economics are the Markov switching models, in which an unobserved discrete state switches stochastically between regimes in which the process evolves in an otherwise linear fashion. Markov switching models were introduced in econometrics by Hamilton (1989) and are also known as hidden Markov models. However, space limitations preclude presenting these models here; for a textbook treatment, see Hamilton (1994). Kim and Nelson (1998, 1999) provide important extensions of this framework to multivariate models with unobserved components. The reader

interested in further discussions of and additional references to other nonlinear time series forecasting methods should see the recent surveys and/or textbook treatments of nonlinear models by Granger and Teräsvirta (1993), Priestly (1989), and Samorodnitsky and Taqqu (1994).

An *artificial neural network* (NN) model relates inputs (lagged values) to outputs (future values) using an index model formulation with nonlinear transformations. There is considerable terminology and interpretation of these formulations which we will not go into here but which are addressed in a number of textbook treatments of these models; see in particular Swanson and White (1995, 1997) for discussions and applications of NN models to economic data. Here, we consider the simplest version, a feedforward NN with a single hidden layer and n hidden units. This has the form:

$$y_{t+h} = \beta_0(L)y_t + \sum_{i=1}^{n} \gamma_i g(\beta_i(L)y_t) + u_{t+h}, \tag{27.6}$$

where $\beta_i(L)$, $i = 0, \ldots, n$ are lag polynomials, γ_i are unknown coefficients, and $g(z)$ is a function that maps $\Re \rightarrow [0, 1]$. Possible choices of $g(z)$ include the indicator function, sigmoids, and the logistic function. A variety of methods are available for the estimation of the unknown parameters of NNs, some specially designed for this problem; a natural estimation method is nonlinear least squares. NNs have a nonparametric interpretation when the number of hidden units (n) is increased as the sample size tends to infinity.

Smooth transition autoregressions are piecewise linear models and have the form:

$$y_{t+h} = \alpha(L)y_t + d_t\beta(L)y_t + u_{t+h}, \tag{27.7}$$

where the mean is suppressed, $\alpha(L)$ and $\beta(L)$ are lag polynomials, and d_t is a nonlinear function of past data that switches between the "regimes" $\alpha(L)$ and $\beta(L)$. Various functions are available for d_t. For example, if d_t is the logistic function so $d_t = 1/(1 + \exp[\gamma_0 + \gamma_1'\zeta_t])$, then the model is referred to as the logistic smooth transition autoregression (LSTAR) model. The switching variable d_t determines the "threshold" at which the series switches, and depends on the data through ζ_t. For example, ζ_t might equal y_{t-k}, where k is some lag for the switch. The parameters of the model can be estimated by nonlinear least squares. Details about formulation, estimation and forecasting for TAR and STAR models can be found in Granger and Teräsvirta (1993) and in Granger, Teräsvirta, and Anderson (1993). For an application of TAR (and other models) to forecasting U.S. unemployment, see Montgomery, Zarnowitz, Tsay, and Tiao (1998).

4.3 Differencing the data

A question that arises in practice is whether to difference the data prior to construction of a forecasting model. This arises in all the models discussed above, but for simplicity it is discussed here in the context of a pure AR model. If one knows a priori that there is in fact a unit autoregressive root, then it is efficient to impose this information and to estimate the model in first differences. Of course,

in practice this is not known. If there is a unit autoregressive root, then estimates of this root (or the coefficients associated with this root) are generally biased towards zero, and conditionally biased forecasts can obtain. However, the order of this bias is $1/T$, so for short horizon forecasts (h fixed) and T sufficiently large, this bias is negligible, so arguably the decision of whether to difference or not is unimportant to first-order asymptotically.

The issue of whether or not to difference the data, or more generally of how to treat the long-term dependence in the series, becomes important when the forecast horizon is long relative to the sample size. Computations in Stock (1996) suggest that these issues can arise even if the ratio, h/T, is small, .1 or greater. Conventional practice is to use a unit root pretest to make the decision about whether to difference or not, and the asymptotic results in Stock (1996) suggest that this approach has some merit when viewed from the perspective of minimizing either the maximum or integrated asymptotic risk, in a sense made precise in that paper. Although Dickey–Fuller (1979) unit root pretests are most common, other unit root tests have greater power, and tests with greater power produce lower risk for the pretest estimator. Unit root tests are surveyed in Stock (1994) and in Chapter 29 in this volume by Bierens.

4.4   Empirical examples

We now turn to applications of some of these forecasting methods to the five US macroeconomic time series in Figures 27.1–27.5.[2] In the previous notation, the series to be forecast, y_t, is the series plotted in those figures, for example, for industrial production $y_t = 200 \ln (\text{IP}_t/\text{IP}_{t-6})$, while for the interest rate y_t is the untransformed interest rate in levels (at an annual rate). The exercise reported here is a simulated out of sample comparison of six different forecasting models. All series are observed monthly with no missing observations. For each series, the initial observation date is 1959:1. Six-month ahead ($h = 6$) recursive forecasts of y_{t+6} are computed for $t = 1971:3, \ldots, 1996:6$; because a simulated out of sample methodology was used, all models were re-estimated at each such date t.

Eight different forecasts are computed: (a) EWMA, where the parameter is estimated by NLS; (b) AR(4) with a constant; (c) AR(4) with a constant and a time trend; (d) AR where the lag length is chosen by BIC ($0 \le p \le 12$) and the decision to difference or not is made using the Elliott–Rothenberg–Stock (1996) unit root pretest; (e) NN with a single hidden layer and two hidden units; (f) NN with two hidden layers, two hidden units in the first layer, and one hidden unit in the second layer; (g) LSTAR in levels with three lags and $\zeta_t = y_t - y_{t-6}$; and (h) LSTAR in differences with three lags and $\zeta_t = y_t - y_{t-6}$.

For each series, the simulated out of sample forecasts (b) and (e) are plotted in Figures 27.1–27.5. The root MSFEs for the different methods, relative to method (b), are presented in Table 27.1; thus method (b) has a relative root MSFE of 1.00 for all series. The final row of Table 27.1 presents the root mean squared forecast error in the native units of the series.

Several findings are evident. First, among the linear models, the AR(4) in levels with a constant performs well. This model dominates the AR(4) in levels with a

Table 27.1   Comparison of simulated out-of-sample linear and nonlinear forecasts for five US macroeconomic time series

Forecasting model		Relative root mean squared forecast errors				
		Unem.	Infl.	Int.	IP	Invent.
(a)	EWMA	1.11	2.55	0.95	1.09	1.50
(b)	AR(4), levels, constant	1.00	1.00	1.00	1.00	1.00
(c)	AR(4), levels, constant and time trend	1.09	1.00	1.05	1.06	1.15
(d)	AR(BIC), unit root pretest	1.05	0.84	1.07	0.99	1.01
(e)	NN, levels, 1 hidden layer, 2 hidden units	1.07	1.76	9.72	1.05	1.54
(f)	NN, levels, 2 hidden layers, 2(1) hidden units	1.07	0.99	0.99	1.07	1.25
(g)	LSTAR, levels, 3 lags, $\zeta_t = y_t - y_{t-6}$	1.04	1.34	3.35	1.02	1.05
(h)	LSTAR, differences, 3 lags, $\zeta_t = y_t - y_{t-6}$	1.04	0.89	1.17	1.01	1.03
Root mean squared forecast error for (b), AR(4), levels, constant		0.61	2.44	1.74	6.22	2.78

Sample period: monthly, 1959:1–1996:12; forecast period: 1971:3–1996:6; forecast horizon: six months.

Entries in the upper row are the root mean squared forecast error of the forecasting model in the indicated row, relative to that of model (b), so the relative root MSFE of model (b) is 1.00. Smaller relative root MSFEs indicate more accurate forecasts in this simulated out-of-sample experiment. The entries in the final row are the root mean squared forecast errors of model (b) in the native units of the series. The series are: the unemployment rate, the six-month rate of CPI inflation, the 90-day US Treasury bill annualized rate of interest, the six-month growth of IP, and the six-month growth of real manufacturing and trade inventories.

Table 27.2   Root mean squared forecast errors of VARs, relative to AR(4)

Forecasting model		Relative mean squared forecast errors				
		Unem.	Infl.	Int.	IP	Invent.
(i)	VAR: $unemp_t$, $int.rate_t$, $\Delta lnIP_t$	0.98	—	1.05	0.93	—
(j)	VAR: $unemp_t$, $\Delta lnCPI_t$, $int.rate_t$	1.03	0.95	1.03	—	—
(k)	VAR: all five variables	1.03	1.10	1.04	0.94	0.96

Sample period: monthly, 1959:1–1996:12; forecast period: 1971:3–1996:6; forecast horizon: six months.

Entries are relative root MSFEs, relative to the root MSFE of model (b) in Table 27.1 (AR(4) with constant in levels). The VAR specifications have lag lengths selected by BIC; the six-month ahead forecasts were computed by iterating one-month ahead forecasts. All forecasts are simulated out of sample. See the notes to Table 27.1.

constant and time trend, in the sense that for all series the AR(4) with a constant and time trend has an RMSFE that is no less than the AR(4) with a constant. Evidently, fitting a linear time trend leads to poor out-of-sample performance, a result that would be expected if the trend is stochastic rather than deterministic. Using BIC lag length selection and a unit root pretest improves upon the AR(4) with a constant for inflation, has essentially the same performance for industrial production and inventories, and exhibits worse performance for the unemployment rate and the interest rate; averaged across series, the RMSFE is 0.99, indicating a slight edge over the AR(4) with a constant on average.

None of the nonlinear models uniformly improve upon the AR(4) with a constant. In fact, two of the nonlinear models ((e) and (g)) are dominated by the AR(4) with a constant. The greatest improvement is by model (h) for inflation; however, this relative RMSFE is still greater than the AR(BIC) forecast for inflation. Interestingly, the very simple EWMA forecast is the best of all forecasts, linear and nonlinear, for the interest rate. For the other series, however, it does not get the correct long-run trend, and the EWMA forecasts are worse than the AR(4).

The final row gives a sense of the performance of these forecasts in absolute terms. The RMSFE of the unemployment rate, six months hence, is only 0.6 percentage points, and the RMSFE for the 90-day Treasury bill rate is 1.7 percentage points. CPI inflation is harder to predict, with a six-month ahead RMSFE of 2.4 percentage points. Inspection of the graph of IP growth reveals that this series is highly volatile, and in absolute terms the forecast error is large, with a six-month ahead RMSFE of 6.2 percentage points.

Some of these points can be verified by inspection of the forecasts plotted in Figures 27.1–27.5. Clearly these forecasts track well the low frequency movements in the unemployment rate, the interest rate, and inflation (although the NN forecast does quite poorly in the 1990s for inflation). Industrial production and inventory growth has a larger high frequency component, which all these models have difficulty predicting (some of this high frequency component is just unpredictable forecast error).

These findings are consistent with the conclusions of the larger forecasting model comparison study in Stock and Watson (1999a). They found that, on average across 215 macroeconomic time series, autoregressive models with BIC lag length determination and a unit root pretest performed well, indeed, outperformed a range of NN and LSTAR models for six-month ahead forecasts. The autoregressive model typically improved significantly on no-change or EWMA models. Thus there is considerable ability to predict many U.S. macroeconomic time series, but much of this predictability is captured by relatively simple linear models with data-dependent determination of the specification.

5   MULTIVARIATE FORECASTS

The motivation for multivariate forecasting is that there is information in multiple economic time series that can be used to improve forecasts of the variable or variables of interest. Economic theory, formal and informal, suggests a

large number of such relations. Multivariate forecasting methods in econometrics are usefully divided into four broad categories: structural econometric models; small linear time series models; small nonlinear time series models; and forecasts based on leading economic indicators.

Structural econometric models attempt to exploit parametric relationships suggested by economic theory to provide a priori restrictions. These models can be hundred-plus equation simultaneous systems, or very simple relations such as an empirical Phillips curve relating changes of inflation to the unemployment rate and supply shocks. Because simultaneous equations are the topic of Chapter 6 by Mariano in this volume, forecasts from simultaneous equations systems will be discussed no further here. Neither will we discuss further nonlinear multivariate models; although the intuitive motivation for these is sound, these typically have many parameters to be estimated and as such often exhibit poor out-of-sample performance (for a study of multivariate NNs, see Swanson and White (1995, 1997); for some positive results, see Montgomery *et al.*, 1998). This chapter therefore briefly reviews multivariate forecasting with small linear time series models, in particular, using VARs, and forecasting with leading indicators. For additional background on VARs, see the chapter in this volume by Lütkepohl.

5.1   Vector autoregressions

Vector autoregressions, which were introduced to econometrics by Sims (1980), have the form:

$$Y_t = \mu_t + A(L)Y_{t-1} + \varepsilon_t, \tag{27.8}$$

where Y_t is a $n \times 1$ vector time series, ε_t is a $n \times 1$ serially uncorrelated disturbance, $A(L)$ is a pth order lag polynomial matrix, and μ_t is a $n \times 1$ vector of deterministic terms (for example, a constant or a constant plus linear time trend). If there are no restrictions on the parameters, the parameters can be estimated asymptotically efficiently (under Gaussianity) by OLS equation by equation. Multistep forecasts can be made either by replacing the left-hand side of (27.8) by Y_{t+h}, or by h-fold iteration of the one-step forecast.

Two important practical questions are the selection of the series to include in Y_t (the choice of n) and the choice of the lag order p in the VAR(p). Given the choice of series, the order p is typically unknown. As in the univariate case, it can be estimated by information criteria. This proceeds as discussed following (27.3), except that $\hat{\sigma}^2$ is replaced by the determinant of $\hat{\Sigma}$ (the MLE of the variance–covariance matrix of ε_t), and the relevant number of parameters is the total free parameters of the VAR; thus, if there are no deterministic terms, $IC(p) = \ln \det(\hat{\Sigma}) + n^2 p g(T)$. The choice of series is typically guided by economic theory, although the predictive least squares (PLS) criterion (which is similar to an information criterion) can be useful in guiding this choice, cf. Wei (1992).

The issue of whether to difference the series is further complicated in the multivariate context by the possible presence of cointegration among two or more of the n variables. The multiple time series Y_t is said to be cointegrated if

each element of Y_t is integrated of order 1 (is I(1); that is, has an autoregressive unit root) but there are $k \geq 1$ linear combination, $\alpha' Y_t$, that are I(0) (that is, which do not have a unit AR root) (Engle and Granger, 1987). It has been conjectured that long-run forecasts are improved by imposing cointegration when it is present. However, even if cointegration is correctly imposed, it remains to estimate the parameters of the cointegrating vector, which are, to first-order, estimated consistently (and at the same rate) if cointegration is not imposed. If cointegration is imposed incorrectly, however, asymptotically biased forecasts with large risks can be produced. At short horizons, these issues are unimportant to first-order asymptotically. By extension of the univariate results that are known for long-horizon forecasting, one might suspect that pretesting for cointegration could improve forecast performance, at least as measured by the asymptotic risk. However, tests for cointegration have very poor finite sample performance (cf. Haug, 1996), so it is far from clear that in practice pretesting for cointegration will improve forecast performance. Although much of the theory in this area has been worked out, work remains on assessing the practical benefits of imposing cointegration for forecasting. For additional discussions of cointegration, see Watson (1994), Hatanaka (1996) and Chapter 30 in this volume by Dolado, Gonzalo, and Marmol.

It should be noted that there are numerous subtle issues involved in the interpretation of and statistical inference for VARs. Watson (1994) surveys these issues, and two excellent advanced references on VARs and related small linear time series models are Lütkepohl (1993) and Reinsel (1993). Also, VARs provide only one framework for multivariate forecasting; for a different perspective to the construction of small linear forecasting models, see Hendry (1995).

5.2   Forecasting with leading economic indicators

Forecasting with leading economic indicators entails drawing upon a large number of time series variables that, by various means, have been ascertained to lead the variable of interest, typically taken to be aggregate output (the business cycle). The first set of leading economic indicators was developed as part of the business cycle research program at the National Bureau of Economic Research, and was published by Mitchell and Burns (1938). More recent works using this general approach include Stock and Watson (1989) and the papers in Moore and Lahiri (1991).

The use of many variables and little theory has the exciting potential to exploit relations not captured in small multivariate time series models. It is, however, particularly susceptible to overfitting within sample. For example, Diebold and Rudebusch (1991) found that although historical values of the Index of Leading Economic Indicators (then maintained by the US Department of Commerce) fits the growth in economic activity well, the real time, unrevised index has limited predictive content for economic activity. This seeming contradiction arises primarily from periodic redefinitions of the index. Their sobering finding underscores the importance of properly understanding the statistical properties of each stage of a model selection exercise. The development of methods for exploiting

large sets of leading indicators without overfitting is an exciting area of ongoing research.

5.3 Empirical examples

We now turn to an illustration of the performance of VARs as forecasting models. Like the experiment reported in Table 27.1, this experiment is simulated out of sample. Three families of VARs were specified. Using the numbering in Table 27.2, model (i) is a three-variable VAR with the unemployment rate, the interest rate, and the growth rate of industrial production. Model (j) is a three-variable VAR with the unemployment rate, CPI inflation, and the interest rate. Model (k) is a VAR with all five variables. The lags in all three models were kept the same in each equation of the VAR and were chosen recursively (at each forecast date, using only data through that date) by BIC, where $1 \leq p \leq 6$ for models (i) and (j), and $1 \leq p \leq 2$ for model (k). The VARs were estimated by OLS, equation by equation, with a one-step ahead specification, and six-month ahead forecasts were computed by iterating the one-month ahead forecasts.

The results are summarized in Table 27.2. In some cases, the VAR forecasts improve upon the AR forecasts. For example, for output growth, the VAR forecasts in (i) and (k) are respectively best and second-best of all the output growth forecasts in both tables. In contrast, for the interest rate, the VAR forecast is worse than the AR(4), and indeed the best forecast for the interest rate remains the EWMA forecast. However, the most notable feature of these forecasts is that eight of the eleven VAR forecasts in Table 27.2 have RMSFEs within 5 percent of the RMSFE of the AR(4), and all eleven have RMSFEs within 10 percent of the AR(4). For these specifications and these series, using additional information via a VAR results in forecasts that are essentially the same as those from an AR(4).

This finding, that forecasts from multivariate models often provide only modest improvements (or no improvement at all) over univariate forecasts, is not new.[3] One way to interpret this result is that additional macroeconomic series have little relationship to one another. This interpretation would, however, be incorrect, indeed, among the relationships in these VARs is the relation between the unemployment rate and inflation (the Phillips curve) and between interest rates and output (a channel of monetary policy), two links that have been studied in great detail and which are robust over the postwar period (cf. Stock and Watson, 1999b). An interpretation more in keeping with this latter evidence is that while these variables are related, there are sufficiently many parameters, which might not be stable over time, that these relations are not particularly useful for multivariate forecasting.

These negative results require some caveats. Supporters of VARs might suggest that the comparison in Table 27.2 is unfair because no attempt has been made to fine tune the VAR, to use additional variables, or to impose prior restrictions or prior information on the lag structure. This criticism has some merit, and methods which impose such structure, in particular Bayesian VARs, have a better track record than the unconstrained VARs reported here; see McNees (1990) and Sims (1993). Alternatively, others would argue that time series models

developed specifically for some variables, such as an empirical Phillips curve (as in Gordon, 1998), would be expected to work better than unfocused application of a VAR. This too might be valid, but in evaluating such claims one must take great care to distinguish between in-sample fit and the much more difficult task of forecasting well out-of-sample, either in real time or in a simulated out-of-sample experiment. Finally, it should be emphasized that these conclusions are for macroeconomic time series. For example, in industry applications one can find series with more pronounced nonlinearities.

6 DISCUSSION AND CONCLUSION

One of the few truly safe predictions is that economic forecasters will remain the target of jokes in public discourse. In part this arises from a lack of understanding that all forecasts must in the end be wrong, and that forecast error is inevitable. Economic forecasters can, however, bolster their credibility by providing information about the possible range of forecast errors. Some consumers are uncomfortable with forecast uncertainty: when his advisors presented a forecast interval for economic growth, President Lyndon Johnson is said to have replied, "ranges are for cattle." Yet communication of forecast uncertainty to those who rely on forecasts helps them to create better, more flexible plans and supports the credibility of forecasters more generally.

A theme of this chapter has been the tradeoff between complex models, which either use more information to forecast or allow subtle nonlinear formulations of the conditional mean, and simple models, which require fitting a small number of parameters and which thereby reduce parameter estimation uncertainty. The empirical results in Tables 27.1 and 27.2 provide a clear illustration of this tradeoff. The short-term interest rate is influenced by expected inflation, monetary policy, and the general supply and demand for funds, and, because the nominal rate must be positive, the "true" model for the interest rate must be nonlinear. Yet, of the autoregressions, neural nets, LSTAR models, and VARs considered in Tables 27.1 and 27.2, the best forecast was generated by a simple exponentially weighted moving average of past values of the interest rate. No attempt has been made to uncover the source of the relatively poor performance of the more sophisticated forecasts of the interest rate, but presumably it arises from a combination of parameter estimation error and temporal instability in the more complicated models.

An important practical question is how to resolve this tradeoff in practice. Two methods have been discussed here. At a formal level, this tradeoff is captured by the use of information criteria. Information criteria can be misleading, however, when many models are being compared and/or when the forecasting environment changes over time. The other method is to perform a simulated out-of-sample forecast comparison of a small number of models. This is in fact closely related to information criteria (Wei, 1992) and shares some of their disadvantages. When applied to at most a few candidate models, however, this has the advantage of providing evidence on recent forecasting performance and how the forecasting performance of a model has evolved over the simulated forecast

period. These observations, along with those above about reporting forecast uncertainty, suggest a simple rule: even if your main interest is in more complicated models, it pays to maintain benchmark forecasts using a simple model with honest forecast standard errors evaluated using a simulated real time experiment, and to convey the forecast uncertainty to the consumer of the forecast.

Finally, an important topic not addressed in this chapter is model instability. All forecasting models, no matter how sophisticated, are stylized and simplified ways to capture the complex and rich relations among economic time series variables. There is no particular reason to believe that these underlying relations are stable – technology, global trade, and macroeconomic policy have all evolved greatly over the past three decades – and even if they were, the implied parameters of the forecasting relations need not be stable. One therefore would expect estimated forecasting models to have parameters that vary over time, and in fact this appears to be the case empirically (Stock and Watson, 1996). Indeed, Clements and Hendry (1999) argue that most if not all major economic forecast failures arise because of unforeseen events that lead to a breakdown of the forecasting model; they survey existing methods and suggest some new techniques for detecting and adjusting to such structural shifts. The question of how best to forecast in a time-varying environment remains an important area of econometric research.

Notes

* The author thanks Lewis Chan for research assistance and four anonymous referees for useful suggestions.
1 All series were obtained from the Basic Economics Database maintained by DRI/McGraw Hill. The series mnemonics are: PUNEW (the CPI); IP (industrial production); LHUR (the unemployment rate); FYGM3 (the 90 day U.S. Treasury bill rate); and IVMTQ (real manufacturing and trade inventories).
2 These results are drawn from the much larger model comparison exercise in Stock and Watson (1999a), to which the reader is referred for additional details on estimation method, model definitions, data sources, etc.
3 In influential work, Cooper (1972) and Nelson (1972) showed this in a particularly dramatic way. They found that simple ARMA models typically produced better forecasts of the major macroeconomic aggregates than did the main large structural macroeconomic models of the time. For a discussion of these papers and the ensuing literature, see Granger and Newbold (1986, ch. 9.4).

References

Bates, J.M., and C.W.J. Granger (1969). The combination of forecasts. *Operations Research Quarterly* 20, 451–68.
Bollerslev, T., R.F. Engle, and D.B. Nelson (1994). ARCH models. In R. Engle and D. McFadden (eds.), *Handbook of Econometrics*, Volume 4, pp. 2959–3038. Amsterdam: Elsevier.
Brockwell, P.J., and R.A. Davis (1987). *Time Series: Theory and Methods*. New York: Springer-Verlag.

Chan, Y.L., J.H. Stock, and M.W. Watson (1998). A dynamic factor model framework for forecast combination. *Spanish Economic Review* 1, 91–121.

Christoffersen, P.F., and F.X. Diebold (1997). Optimal prediction under asymmetric loss. *Econometric Theory* 13, 808–17.

Clements, M., and D.F. Hendry (1996). Multi-step estimation for forecasting. *Oxford Bulletin of Economics and Statistics* 58, 657–84.

Clements, M., and D.F. Hendry (1999). *Forecasting Non-Stationary Economic Time Series*. Cambridge: MIT Press.

Cooper, R.L. (1972). The predictive performance of quarterly econometric models of the United States. In B.G. Hickman (ed.) *Econometric Models of Cyclical Behavior*. New York: Columbia University Press.

Davis, R.A., and W.T.M. Dunsmuir (1996). Maximum likelihood estimation for MA(1) processes with a root on or near the unit circle. *Econometric Theory* 12, 1–29.

Diebold, F.X., T. Gunther, and A.S. Tay (1998). Evaluating density forecasts with applications to financial risk management. *International Economic Review* 39, 868–83.

Diebold, F.X., and J.A. Lopez (1995). Forecast evaluation and combination. In G.S. Maddala and C.R. Rao (eds.) *Handbook of Statistics* 14, 241–68.

Diebold, F.X., and R. Mariano (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics* 13, 253–63.

Diebold, F.X., and G. Rudebusch (1991). Forecasting output with the composite leading index: a real time analysis. *Journal of the American Statistical Association* 86, 603–10.

Elliott, G., T.J. Rothenberg, and J.H. Stock (1996). Efficient tests for an autoregressive unit root. *Econometrica* 64, 813–36.

Engle, R.F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation. *Econometrica* 50, 987–1008.

Engle, R.F., and C.W.J. Granger (1987). Co-integration and error correction: representation, estimation and testing. *Econometrica* 55, 251–76.

Fair, R.C. (1980). Evaluating the predictive accuracy of econometric models. *International Economic Review* 21, 355–78.

Geweke, J., and R. Meese (1981). Estimating regression models of finite but unknown order. *International Economic Review* 22, 55–70.

Gordon, R.J. (1998). Foundations of the Goldilocks economy: supply shocks and the time-varying NAIRU. *Brookings Papers on Economic Activity* 2, 297–333.

Granger, C.W.J. (1969). Prediction with a generalized cost of error function. *Operational Research Quarterly* 20, 199–207.

Granger, C.W.J. (1989). Combining forecasts – twenty years later. *Journal of Forecasting* 8, 167–73.

Granger, C.W.J., and P. Newbold (1986). *Forecasting Economic Time Series*, 2nd edn. Orlando: Academic Press.

Granger, C.W.J., and T. Teräsvirta (1993). *Modelling Non-linear Economic Relationships*. Oxford: Oxford University Press.

Granger, C.W.J., T. Teräsvirta, and H.M. Anderson (1993). Modeling nonlinearity over the business cycle. In J.H. Stock and M.W. Watson (eds.) *Business Cycles, Indicators, and Forecasting*, pp. 311–27. Chicago: University of Chicago Press for the NBER.

Hamilton, J.D. (1994). *Time Series Analysis*. Princeton: Princeton University Press.

Hamilton, J.D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica* 57, 357–84.

Harvey, A.C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge: Cambridge University Press.

Hatanaka, M. (1996). *Time-Series-Based Econometrics: Unit Roots and Cointegration*. Oxford: Oxford University Press.

Haug, A.A. (1996). Tests for cointegration: a Monte Carlo comparison. *Journal of Econometrics* 71, 89–115.

Hendry, D.F. (1995). *Dynamic Econometrics*. Oxford: Oxford University Press.

Kim, C.-J., and C.R. Nelson (1998). Business cycle turning points, a new coincident index, and tests of duration dependence based on a dynamic factor model with regime switching. *Review of Economics and Statistics* 80, 188–201.

Kim, C.-J., and C.R. Nelson (1999). *State-Space Models with Regime Switching: Classical and Gibbs Sampling Approaches with Applications*. Cambridge: MIT Press.

Lütkepohl, H. (1993). *Introduction to Multiple Time Series Analysis*, 2nd edn. New York: Springer-Verlag.

McNees, S.K. (1990). The role of judgment in macroeconomic forecasting accuracy. *International Journal of Forecasting* 6, 287–99.

Meese, R., and J. Geweke (1984). A comparison of autoregressive univariate forecasting procedures for macroeconomic time series. *Journal of Business and Economic Statistics* 2, 191–200.

Mitchell, W.C., and A.F. Burns (1938). *Statistical Indicators of Cyclical Revivals*. NBER Bulletin 69, New York. Reprinted in G.H. Moore (ed.) *Business Cycle Indicators*. Princeton: Princeton University Press, 1961.

Moore, G., and K. Lahiri (eds.) (1991). *The Leading Economic Indicators: New Approaches and Forecasting Records*. Cambridge: Cambridge University Press.

Montgomery, A.L., V. Zarnowitz, R.S. Tsay, and G.C. Tiao (1998). Forecasting the U.S. unemployment rate. *Journal of the American Statistical Association* 93, 478–93.

Nelson, C.R. (1972). The prediction performance of the FRB-MIT-PENN model of the U.S. economy. *American Economic Review* 62, 902–17.

Ng, S., and P. Perron (1995). Unit root tests in ARMA models with data dependent methods for the truncation lag. *Journal of the American Statistical Association* 90, 268–81.

Priestly, M.B. (1989). *Non-linear and Non-stationary Time Series Analysis*. London: Academic Press.

Reinsel, G.C. (1993). *Elements of Multivariate Time Series Analysis*. New York: Springer-Verlag.

Samorodnitsky, G., and M.S. Taqqu (1994). *Stable Non-Gaussian Random Processes*. New York: Chapman & Hall.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics* 6, 461–64.

Sims, C.A. (1980). Macroeconomics and reality. *Econometrica* 48, 1–48.

Sims, C.A. (1993). A nine-variable probabilistic macroeconomic forecasting model. In J.H. Stock and M.W. Watson (eds.) *Business Cycles, Indicators and Forecasting*. Chicago: University of Chicago Press for the NBER.

Stock, J.H. (1994). Unit roots, structural breaks, and trends. In R. Engle and D. McFadden (eds.) *Handbook of Econometrics*, Volume 4, pp. 2740–843. Amsterdam: Elsevier.

Stock, J.H. (1996). VAR, error correction and pretest forecasts at long horizons, *Oxford Bulletin of Economics and Statistics* 58, 685–701. Reprinted in A. Banerjee and D.F. Hendry (eds.) *The Econometrics of Economic Policy*, pp. 115–32. Oxford: Basil Blackwell, 1997.

Stock, J.H., and M.W. Watson (1989). New indexes of coincident and leading economic indicators. *NBER Macroeconomics Annual* 351–93.

Stock, J.H., and M.W. Watson (1996). Evidence on structural instability in macroeconomic time series relations. *Journal of Business and Economic Statistics*, 14, 11–30.

Stock, J.H., and M.W. Watson (1998). Median unbiased estimation of coefficient variance in a time varying parameter model. *Journal of the American Statistical Association* 93, 349–58.

Stock, J.H., and M.W. Watson (1999a). A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series. Chapter 1 in R. Engle and H. White

(eds.) *Cointegration, Causality and Forecasting: A Festschrift for C.W.J. Granger*. Oxford: Oxford University Press, 1–44.

Stock, J.H., and M.W. Watson (1999b). Business cycle fluctuations in U.S. macroeconomic time series. Chapter 1 in J. Taylor and M. Woodford (eds.) *Handbook of Macroeconomics*. Amsterdam: Elsevier, 3–64.

Swanson, N.R., and H. White (1995). A model selection approach to assessing the information in the term structure using linear models and artificial neural networks. *Journal of Business and Economic Statistics* 13, 265–75.

Swanson, N.R., and H. White (1997). A model selection approach to real-time macroeconomic forecasting using linear models and artificial neural networks. *Review of Economics and Statistics* 79, 540–50.

Watson, M.W. (1994). Vector autoregressions and cointegration. In R. Engle and D. McFadden (eds.) *Handbook of Econometrics*, Volume 4, pp. 2844–915. Amsterdam: Elsevier.

Wei, C.Z. (1992). On predictive least squares principles. *The Annals of Statistics* 20, 1–42.

West, K. (1996). Asymptotic inference about predictive ability. *Econometrica* 64, 1067–84.

West, K., H.J. Edison, and D. Cho (1993). A utility based evaluation of some models of exchange rate variability. *Journal of International Economics* 35, 23–46.

Zarnowitz, V., and Braun (1993). Twenty-two years of the NBER-ASA quarterly economic outlook surveys: aspects and comparisons of forecasting performance. In J.H. Stock and M.W. Watson (eds.) *Business Cycles, Indicators and Forecasting*. Chicago: University of Chicago Press for the NBER.