Journal of Economic Literature 2008, 46:1, 3–56
http:www.aeaweb.org/articles.php?doi=10.1257/jel.46.1.3

Economic Forecasting

GRAHAM ELLIOTT AND ALLAN TIMMERMANN*

Forecasts guide decisions in all areas of economics and finance and their value can only be understood in relation to, and in the context of, such decisions. We discuss the central role of the loss function in helping determine the forecaster's objectives. Decision theory provides a framework for both the construction and evaluation of forecasts. This framework allows an understanding of the challenges that arise from the explosion in the sheer volume of predictor variables under consideration and the forecaster's ability to entertain an endless array of forecasting models and time-varying specifications, none of which may coincide with the "true" model. We show this along with reviewing methods for comparing the forecasting performance of pairs of models or evaluating the ability of the best of many models to beat a benchmark specification.

1. *Introduction*

Forecasting problems are ubiquitous in all areas of economics and finance where agents' decisions depend on the uncertain future value of one or more variables of interest. When a household decides how much labor to supply or how much to save for a rainy day, this presumes an ability to forecast a stream of future wages and returns on savings. Similarly, firms' choice of when to invest, how much to invest, and how to finance it (the capital structure decision) depends on their forecasts of future cash flows from potential investments, future stock prices, and interest rates. Indeed, all present value calculations, and hence the vast majority of questions in asset pricing, have embedded in them forecasts of future cash flows generated by uncertain payoff streams. In public finance, decisions on whether to go ahead with large infrastructure projects, such as the construction of a new bridge or a tunnel, require projecting traffic flows and income streams over the project's lifetime, which may well be several decades.

Recent research has seen a virtual revolution in how economists compute, apply, and evaluate forecasts. This research has occurred as a result of extensive developments in information technology that have opened access to thousands of new potential predictor variables (including tick-by-tick trading data, disaggregate survey forecasts, and real-time macroeconomic data) and a wealth of new techniques that facilitate search over and estimation of the parameters of increasingly complicated forecasting models. Questions such as which particular predictor variables

* Elliott: University of Californa, San Diego. Timmermann: University of Californa, San Diego, and CREATES.
We thank the editor, Roger Gordon, three anonymous referees and Lutz Kilian, Michael McCracken, Barbara Rossi, and Norm Swanson for providing detailed comments on the paper. We also thank Gray Calhoun for excellent research assistance. Timmermann acknowledges research support from CREATES funded by the Danish National Research Foundation.

to include, which functional form to use for the forecasting model, and how to weight old versus more recent data have become an essential part of forecast construction and evaluation.

Economic forecasting is unique in that forecasters are forced to "show their hand" in real time as they generate their forecasts. Future outcomes of most predicted variables are observed within a reasonable period of time, so a direct sense of how well a forecasting model performed can be gained. If forecasting performance is poor, this will become clear to the forecaster once data on realizations of the predicted variable is revealed. This real time feedback may in turn lead to a change in the forecasting model itself, thus posing unique challenges to the process of evaluating how fast the forecaster is learning over time. This is in stark contrast to many econometric problems. For example, evaluation of an estimate of the effect of schooling on wages may take years. In many economic problems, we do not obtain an objective confirmation of how good the original estimate is since we do not have reference data for evaluating the economic prediction.

Often the result of the feedback from forecasts has been disheartening, both to econometricians trying to utilize data as efficiently as possible and to economists whose theories result in predictions that appear unable to explain as much of the variation in the data as they had hoped. To take one example, a seemingly simple task such as estimating the weights on different models in least squares forecast combination regressions is commonly outperformed on real data by using a simple equal-weighted average of forecasts (Robert T. Clemen 1989). An infamous result of Richard A. Meese and Kenneth Rogoff (1983) shows that despite a great deal of theoretical work on exchange rates—and even with the benefit of using future data suggested by theory as relevant—the random walk "no change" prediction cannot be beaten. This result has to a great extent held up for exchange rate forecasts (Lutz Kilian 1999).

While the performance of a forecasting model often can be observed fairly quickly, only limited economic conclusions can be drawn from the model's historical track record. Forecasting models are best viewed as greatly simplified approximations of a far more complicated reality and need not reflect causal relations between economic variables. Indeed, simple mechanical forecasting schemes—such as the random walk—are often found to perform well empirically although they do not provide new economic insights into the underlying variable (Michael P. Clements and David F. Hendry 2002). Conversely, models aimed at uncovering true unconditional relationships in the data need not be well suited for forecasting purposes.

In a unified framework, this paper provides an understanding of the properties, construction, and evaluation of economic forecasts. Our objective is to help explain differences among the many approaches used by various researchers and understand the breadth of results reported in the empirical forecasting literature.

Our coverage emphasizes the importance of integrating economic forecasts (including model specification, variable selection, and parameter estimation) in a decision theoretical framework. This is, in our view, the defining characteristic of economic forecasts. From this perspective, forecasts do not have any intrinsic value and are only useful in so far as they help improve economic decisions. What constitutes a good forecast depends on how costly various prediction errors are to the forecaster and hence reflects both the forecaster's preferences and the manner in which forecasts are mapped into economic decisions. Economic forecasting is not an exercise in modeling the data disjoint from the purpose of the forecast provision. Section 2 illustrates these points initially through two examples from economics and finance.

We next provide a formal statement of the forecasting problem. Section 3 reviews both classical and Bayesian approaches to

forecasting and introduces the individual components of the economic forecasting problem such as the forecaster's prediction model, the underlying information set as well as the forecaster's loss function. Although often only treated implicitly, the loss function is essential to all forecasting problems and so we devote section 4 to a deeper discussion of various types of loss functions and the restrictions and assumptions they embody.

Using the decision theoretic framework set out in section 3, sections 5–7 review several issues that arise in the practical construction of economic forecasts. Each of these topics has been active areas of research in recent years. An overarching problem in economic forecasting is the myriad of data that a forecaster could potentially employ. Estimating models with a large number of parameters relative to the sample size undermines one of the central methods of econometrics—ordinary least squares (OLS) justified through properties such as asymptotic efficiency of the parameter estimates—and opens the possibility that other estimation techniques are better suited to the task of constructing forecasts. In concert with differences over loss functions, this provides a partial explanation of the myriad of estimation methods seen in practice.

Section 5 addresses the choice of functional form of the forecasting model. Lack of guidance from economic theory is often an issue and so the functional form is commonly chosen on grounds such as empirical "fit" or an ability to capture certain episodes in the historical data sample. We review several methods aimed at approximating unknown functional forms in a parsimonious, yet flexible manner—a task made essential by the short samples available in most forecasting applications.

A related problem is how the forecasting model and the underlying data evolve over time. When forecasting models are viewed as simple approximations to a complex and evolving reality that changes due to shifts in legislation, institutions, and technology—or even wars and natural catastrophes—it is to be expected that the "true" but unknown data generating process changes over time. In the forecasting literature, this has been captured through various approaches that deal with model and parameter instability. Since all estimation techniques essentially average over past data to obtain a forecasting model, this raises the problem of exactly how to choose the data sample and how to weight "old" versus "new" data. Other approaches attempt to directly model breaks in the model parameters in order to increase the effective data sample. These are covered in section 6.

An important part of the analysis of economic forecasts is to assess how good they are. Until recently, forecasts were largely evaluated without the use of standard errors that account for parameter estimation error and model specification search. It is well understood that data mining programs that search over many models tend to overfit and hence inflate estimates of forecasting performance by using the same data for estimation and evaluation purposes. However, little or no account is typically made for such model search that precede the analysis. Standard practice for dealing with data mining has been to hold back some data and check whether the forecasting model still performed well in future out-of-sample periods. Again, often average losses are compared without any regard to pretesting biases. Recent work has resulted in methods that account for sampling error in various forecasting situations. These are reviewed in section 7.

Economic theory rarely identifies a single forecasting model that works well in practice and leaves open many degrees of freedom in forecast construction. The resulting plethora of economic forecasting models has given rise to procedures for forecast comparisons that can handle situations with a very large set of models. An alternative to evaluating particular models and attempting to select a single dominant model is to average over various forecasting methods. Both forecast

comparison and forecast combination are reviewed in section 8. Section 9 provides an empirical analysis of forecasts of inflation and stock returns. Finally, section 10 concludes.

2. *Forecasts and Economic Decisions: Two Examples*

We start with an illustration of how economic forecasts are embedded in the economic decision process using two examples from macroeconomics and finance.

2.1 *Central Bank Forecasts*

Consider the forecasting problem encountered by a central bank whose main role is to set interest rates and whose objectives are defined over inflation and economic activity as measured, e.g., by output growth and the unemployment rate. Because future values of output growth, unemployment, and inflation are uncertain, in practice the bank's interest rate decisions depend on its forecast of these variables as well as its understanding of how they will be affected by current and future interest rates. In the analysis by Lars E. O. Svensson (1997), the central bank's inflation targeting implies targeting inflation forecasts and so the forecast effectively acts as an intermediate target.

As part of formulating a forecasting model, the central bank must decide which variables are helpful in predicting future economic activity and inflation. Forecasts of output growth and inflation could be linked via the Phillips curve. Monetary theories of inflation may suggest one set of variables, while the theory of the term structure of interest rates may suggest others. Variables such as new housing starts, automobile sales, new credit lines, monetary growth, personal bankruptcies, capacity utilization, unemployment rates, etc. must also be assessed. Even after determining which predictor variables to include in the forecasting model, questions such as how to measure a particular variable or which dynamic lag structure to use must also be addressed.

When more than one forecasting model is available, which model to use—or whether to use a combination of forecasts from separate models—also becomes an issue. Many central banks make use of what Adrian Pagan (2003) refers to as a diverse "suite of models." Indeed, according to Pagan, at some stage the Bank of England made use of thirty-two different models (although not all of these were used in forecasting), ranging from VARs, time-varying component models to factor models. Similar evidence on the use of multiple models by other central banks is reported by Christopher A. Sims (2002).

Because central banks' quantitative models serve the dual purposes of being used in policy analysis and forecasting, a trade-off is likely to exist between the models' theoretical and empirical coherence (Pagan 2003). For example, theory may impose constraints on the behavior of equilibrium error correction mechanisms such as the gradual disappearance of the output gap. The existence of such a trade-off means that the central bank may choose not to maximize the pure statistical "fit" of the forecasting model when this is deemed to compromise the model's theoretical coherence.

Closely related to this point, a key characteristic of forecasts produced by central banks is that they are often conditional forecasts computed for a prespecified path of future interest rates. Such conditional forecasts are usually computed in the context of a structural model for the economy. Since current and future interest rates are affected by the central bank's decisions, the central bank's forecasting problem cannot be separated from its decisions regarding current and future interest rates.

Central bankers come with certain subjective views about how the economy operates that they may wish to impose on their forecasting model—a theme that naturally leads to Bayesian forecasting methods or other methods that can trade off theoretical and empirical coherence. Should the central bank use a simple vector autoregression (VAR)

fitted to historical data or maybe as a way to capture expectations as was done at least at some point by the FRB/US model used at the Board (F. Brayton and P. Tinsley 1996) and thus use a model tailored to fit historical features of the data? Should it use a more theoretically coherent dynamic stochastic general equilibrium (DSGE) model? Or, should it use some combination of the two? If the central bank adjusts forecasts from a formal model using judgmental information, an additional issue arises, namely how much weight to assign to the data versus the judgmental forecast. Implicitly or explicitly, such weights will reflect the bank's prior beliefs.

Model instability or "breaks" are likely to be empirically relevant for central bankers trying to forecast inflation. In fact, inflation appears to be among the least stable macroeconomic variables exactly because it depends on monetary policy regimes, macroeconomic shocks, and other factors. James H. Stock and Mark W. Watson (1999b) report evidence of instability in the parameters of the Phillips curve.

Quite frequently forecasters find themselves in situations that differ in important regards from the historical sample used to estimate their forecasting models. Pagan (2003) refers to the difficulties and uncertainties the Bank of England faced in their forecasts following the events of September 11, 2001. Indeed, an important part of maintaining a good forecasting model is to monitor and evaluate its performance both historically and in real time. Because past forecast errors have often been found to have predictive power over future errors, monitoring for serial correlation in forecast errors potentially offers a simple way to improve upon a forecast. More generally, if the process generating the predicted variable is subject to change, it is conceivable that a forecasting model that performed well historically may have failed to do so in the more recent past.

2.2 *Portfolio Allocation Decisions*

As a second example, consider an investor's portfolio allocation decisions. Under mean–variance preferences, these will depend on the investor's forecasts of a set of assets' mean returns as well as their variances and covariances. Under more general preferences, higher order moments such as skew and kurtosis and possibly the full return distribution may also matter to the investor. In either case, the investor must be able to produce quantitative forecasts and trade off portfolios with different probability distributions through a loss function. The investor must also decide how to incorporate predictability into his actions. How predictability maps into portfolio allocations will depend on the form of the prediction signal—i.e., is the sign of asset returns predictable or only their magnitude?

Even if conditional means and variances of asset returns are believed to be constant and hence essentially unpredictable, their estimates can still be surrounded by considerable uncertainty. As a consequence, how the moments are estimated can in practice have a large effect on the portfolio weights. Due to estimation error, often the raw estimates are shrunk toward their values implied by a simple benchmark model such as the capital asset pricing model (Olivier Ledoit and Michael Wolf 2003). Alternatively, the investor's choice variables—the portfolio weights—can be restricted through short sale restrictions and maximum holding limits (Ravi Jagannathan and Tongshu Ma 2003).

If the mean and variance of returns are allowed to depend on time-varying state variables, the question immediately arises which state variables to select among interest rates (levels and spreads), macroeconomic activity variables, technical variables such as price momentum or reversals, valuation measures such as the price–earnings or book-to-market ratios or the dividend yield etc. Asset pricing theory provides little guidance to the exact identity of the relevant state variables; this raises several questions such as how to avoid overfitting the forecasting model—a risk always encountered when multiple prediction models are considered—and how to assess the forecasting models' performance

against a benchmark strategy such as simply holding the market portfolio.

Another problem that is more unique to forecasting models for financial returns is that any predictability patterns that do not capture time-varying risk premia must, if markets are efficient, be nonstationary because their discovery should lead to their self-destruction once investors act to take advantage of such predictability. For example, there is evidence suggesting that popular models for predicting stock returns based on the dividend yield ceased to be successful at some point during the 1990s, perhaps because of changes in firms' dividend payout and share repurchase practices or perhaps because investors incorporated earlier evidence of predictability. Only if a model's forecasting performance is tracked carefully through time can this sort of evidence be uncovered.

These examples indicate the complexity of many of the issues involved in economic forecasting. To further understand these points, we next provide a formal statement of the objectives underlying the calculation of actual forecasts.

3. *A Formal Statement of the Forecasting Problem*

Forecasting can be broadly viewed as the process involved in providing information on future values of one or more variables of interest. Toward this end, the variables of interest must be defined and the information set containing known data that will be considered to construct the forecast must also be determined. The latter can be problematic in practice since we often have very large amounts of information that could be used as inputs to the forecasting model.

Other elements are important to the process of deriving a forecast, some of which are often ignored to some extent even though implicitly they still play a role.

The first element is the loss function. No forecast is going to always be correct, so a specification of how costly different mistakes are is needed to guide the procedure. This helps to avoid—or at least lower the probability of—worst case scenarios.

The second element is the family of forecasting models to be considered. This guides the selection of possible methods used for forecast construction. Models may be parametric, semiparametric or nonparametric. A parametric model is a model which is fully specified up to a finite dimensional unknown vector. A nonparametric model can be considered as a model with an infinite dimensional set of unknown parameters. Semi-parametric models fit in the middle. Since little is often known about the form of the "true" forecasting model, ideally one would specify the forecasting model nonparametrically. However, this ignores the short data samples and the large dimension of the set of potential predictor variables in most empirical forecasting problems. In practice, a flexible parametric forecasting model is often the best one can hope to achieve.

A third element concerns which type of information to report for the outcome of interest. We could report a single number (point estimate), a range estimate, or perhaps an estimate of the full probability distribution of all possible values. Most of the theory of frequentist forecasting has been directed toward point forecasting. A more recent literature has examined interval forecasts or forecasts of the conditional distribution of the variable of interest rather than a summary statistic. From a Bayesian perspective, similar issues arise, although it is natural in this approach to provide the full predictive distribution.

3.1 *Notation*

Throughout the analysis, we let Y be the random variable that generates the value to be forecast. To begin with, we restrict attention to point forecasts, f, which are functions of the available data at the time the forecast is made. Hence, if we collect all relevant information at the time of the forecast into the outcome z of the random variable Z, then the forecast is $f(z)$. Discovering which

variables are informative from a forecasting standpoint is important in practice. We examine this in greater depth later on but for now simply think of this information as being incorporated into some random variable that generates our data. Exactly how z maps into the forecast $f(z)$ depends on a set of unknown parameters, θ, that typically have to be estimated from data. We emphasize this by writing $f(z,\theta)$.

The loss function is a function $\mathcal{L}(f,Y,Z)$ that maps the data, Z, outcome, Y, and forecast, f, to the real number line, i.e., for any set of values for these random variables the loss function returns a single number. The loss function describes in relative terms how bad any forecast might be given the outcome and possibly other observed data that accounts for any state dependence in the loss. The loss function and its properties are examined in greater detail in section 4.

3.2 *Optimal Point Forecasts*

The forecaster's objective is to use data—outcomes of the random variable Z—to predict the value of the random variable Y. Let T be the date where the forecast is computed and let h be the forecast horizon. Then Z is defined on the information set $\mathcal{F}_T$, while the outcome is defined on $\mathcal{F}_{T+h}$. Z comprises a sequence $\{Z_t\}_{t=1}^T$ that typically includes current and past values of the variable to be forecast, as well as other variables, X, so often $\{Z\}_{t=1}^T = \{Y_t, X_t\}_{t=1}^T$. The outcome, Y, may be a vector or could be univariate.

The forecaster's objective can be reduced to finding a decision rule $f(z)$ that will be used to choose a value for the outcome of Y. The forecast is the decision rule. For any decision rule, there is an attached "risk,"

$$(1) \qquad R(\theta,f) = E_{Y,Z}[\mathcal{L}(f(Z,\theta),Y,Z)].$$

Here the expectation is over the data Z and the outcome Y holding the forecasting rule f and the unknown parameters, θ, fixed (which is why the risk is a function of θ and the particular rule chosen, f). That this is a function of θ will become clear below. A sensible rule has low risk and minimizes expected loss or equivalently maximizes expected utility.[1]

Assuming the existence of a density for both Y given Z and for Z (denoted $p_{Y}(y|z,\theta)$ and $p_{Z}(z|\theta)$, respectively), we can write the risk as

$$(2) \qquad R(\theta,f) = \int_z \int_y \mathcal{L}(f(z,\theta),y,z)p_{Y}(y|z,\theta)p_{Z}(z|\theta)\,dydz.$$

It is this risk that forecasting methods—methods for choosing $f(z,\theta)$—attempt to control and minimize.

Forecasts are generally viewed as "poor" if they are far from the observed realization of the outcome variable. However, as is clear from (2), point forecasts aim to estimate not the realization of the outcome but rather a function of its distribution. The inner integral in the risk function (2) is $E_{Y}[\mathcal{L}(f(z,\theta),Y,Z)|Z]$, which removes Y from the expression leaving the loss function relating the forecast to θ and the realization of the data z. For example, in the case of mean squared loss, this is the variance of Y given Z plus the squared difference between the forecast and the mean of Y given Z. Both the conditional mean and variance of Y are functions of θ and z.

As noted in section 2.1, we are often interested in conditional forecasts, i.e., forecasts of Y conditional on a specific path taken by another random variable, W. In the above analysis and in what follows, the results can be extended to this case by replacing $p_{Y}(y|z,\theta)$ and $p_{Z}(z|\theta)$ with the distributions conditional on the outcome of W being set to w, i.e., $p_{Y}(y|z,w,\theta)$ and $p_{Z}(z|w,\theta)$. While this extension may seem trivial conceptually, it

[1] This representation of the problem limits further choices of the loss function and requires assumptions on the underlying random variables to ensure that the risk exists.

can be difficult to implement in practice. For example, consider a VAR in interest rates and inflation, where we want to predict inflation one period ahead conditional on the value of interest rates one period ahead. The density of future inflation given future interest rates as well as current and past values of both variables is relatively simple to write down or estimate and is the density of a rotated VAR. However, the density of past values of both variables conditional on future interest rates presents some difficulties in practice (see, e.g., Daniel F. Waggoner and Tao Zha 1999 for a Bayesian example).

3.3 *Classical Approach*

The classical approach to forecasting focuses on evaluating the inner integral $\int \mathcal{L}(f(z,\theta),y,z)p_Y(y|z,\theta)\,dy$. This expectation is taken with respect to the outcome variable holding both the data used to construct the forecast, z, and the parameters, θ, fixed. For a given conditional density for Y (i.e., a model for Y conditional on $Z = z, p_Y(y|z,\theta)$), a set of parameters, θ, and a given loss function, $\mathcal{L}$, we can minimize this directly for a rule $f(z,\theta)$.

If we are able to differentiate under the integral, we get the forecaster's first order condition

$$(3)\quad \frac{d}{df}\int_y \mathcal{L}(f(z,\theta),y,z)p_Y(y|z,\theta)\,dy = E[\mathcal{L}'(f(Z,\theta),Y,Z)|Z=z,\theta] = 0,$$

where $\mathcal{L}'$ (evaluated using the data) is often called the generalized forecast error. Assuming squared loss in the difference between the forecast and outcome, we have

$$(4)\quad \int_y \mathcal{L}(f(z,\theta),y,z)p_y(y|z,\theta)\,dy = E[Y - f(Z,\theta)|Z=z,\theta]^2.$$

This is minimized by choosing $f(z,\theta) = E[Y|Z=z,\theta]$, i.e., the conditional mean as a function of the data, Z, the forecasting model, and its parameters, θ.

In practice, the parameters θ are almost always unknown and so the second step in the classical approach involves selecting a "plug-in" estimator for θ. The resulting estimator $\hat{\theta}(z)$ is a function of the data z and hence the forecasting rule $f(z,\hat{\theta}(z))$ is only a function of the observable data.

For example, under squared loss and $z = \{y_t,x_t\}_{t=1}^T$, when the conditional mean of Y_{T+1} is $\theta' x_T$ we might use OLS estimates from a regression of y_{t+1} on x_t over the available sample as the plug-in estimator for θ. The forecast is then $f(z,\hat{\theta}(z)) = (\sum_{t=2}^T x'_{t-1}y_t)'(\sum_{t=2}^T x'_{t-1}x_{t-1})^{-1}x_T$. Alternative plug-in estimators are discussed in detail below.

In choosing between plug-in estimators, one approach is to examine the risk functions for the various methods, $R(\theta,f)$. These are functions of both the method f and the parameters θ. Typically no risk function dominates uniformly over all θ, i.e., some are better for some values of θ but work less well for other values. The classical forecaster could then choose a method that minimizes worst case risk or could alternatively consider a weighting function over θ, choosing the best method for that particular weighting. Denoting the weighting function by $\pi(\theta)$, one would choose the method that minimizes $\int R(\theta,f)\pi(\theta)\,d\theta$, i.e., the risk averaged over all models that are thought to be important.

3.4 *Bayesian Approach*

The Bayesian approach starts with the idea of risk averaged over all possible models by defining Bayes risk as

$$(5)\quad r(\pi,f) = \int_\theta R(\theta,f)\pi(\theta)\,d\theta.$$

If the forecast $f(z,\theta)$ minimizes Bayes risk, it is a Bayes decision rule. Notice the similarity

to choosing the plug-in method that minimizes average risk over relevant models in the classical approach.

To construct a Bayes decision rule (i.e., a forecast) we require a weighting or prior, $\pi(\theta)$, over the parameters of the model that tells us which parameter values are likely and which are not. We also require models for the random variables underlying the data, namely $p_Y(y|z,\theta)$ and $p_Z(z|\theta)$, which allow calculation of the posterior $\pi(\theta|z) = p_Z(z|\theta)\pi(\theta)/m(z)$, where $m(z) = \int p_Z(z|\theta)\pi(\theta)\,d\theta$. Bayesian forecasts are then chosen conditional on observing $Z = z$, using the posterior.

To see intuitively why this works in the sense of delivering a forecast that minimizes Bayes risk, consider expected loss conditional on $Z = z$. A Bayesian approach would be to minimize this for any $Z = z$, yielding a rule $f(z,\theta)$. Since this rule minimizes the conditional expected loss, it also minimizes the unconditional expected loss, i.e., it minimizes Bayes risk and is hence a Bayesian decision rule.[2]

3.5 *Relating the Methods*

The complete class theorem tells us that if the classical method does not correspond to a Bayesian procedure for some prior, then it is inadmissible. Under the same problem setting (i.e., for identical loss function and densities), one could therefore find a Bayesian method with equal or smaller risk than the classical procedure for all possible values of θ. Conversely, if there is equivalence between the two methods, then the classical approach cannot be beaten. Admissibility is only interesting for weights $\pi(\theta)$ relevant to the forecasting problem, but will of course be preferred given such weights.

The first problem that can arise in the classical setting is that the "plug in" method of constructing $f(z,\theta)$ using the estimate $\hat{\theta}(z)$ is ad hoc. Often forecasters choose estimators that yield nice properties of the parameters themselves. For example, estimators that are consistent, asymptotically normal, and asymptotically efficient for θ may be employed. However, because the goal of forecasting is not to estimate θ, but to construct the forecast, $f(z,\theta)$, such methods may not yield good forecasting rules even for reasonable weighting functions $\pi(\theta)$.

Some practical considerations cloud the picture. In practice, differences between the risks of optimal and ad hoc methods need not be large enough to justify differences in computational costs. Moreover, such comparisons require that the model be correctly specified which is against both the spirit and practice of modern forecasting. Still, the Bayesian approach offers a construction method that is guaranteed to be admissible for the specified model. Even if the true model is not necessarily the one used to construct the forecasts, provided that the forecasting model is close to the true model we are assured to use a method that works well for this possible true model.

It follows from this discussion that it is difficult to find optimal solutions even for very simple forecasting problems. Furthermore, even for simple models such as linear regression models OLS may not be the best approach (this is discussed at length in section 5.1). Hence for various combinations of distributions of the data and values of the parameters of the model there is leeway for alternative methods to dominate. This lack of a single dominant approach explains much of the interest in different forecasting approaches seen in the last two decades.

3.6 *Density Forecasts*

An alternative to the provision of a point forecast is to provide an estimate of the predictive density $p_Y(y|z)$, usually termed the density forecast. Knowledge of this density is sufficient for finding a solution to minimizing $E_Y[\mathcal{L}(f(z,\theta),Y,Z)|Z]$ over decision rules, i.e.,

[2] John Geweke (2005) provides an extensive treatment of Bayesian methods.

some function of the density forecast, that minimizes risk

$$(6)\quad f^*(z,\theta) = \arg\min_{f(z,\theta)} \int_y \mathcal{L}(f(z,\theta),y,z)p_Y(y|z)\,dy.$$

Closed form solutions are not always available, so numerical integration over the density forecasts is often required to evaluate the risk.

Forecasters with different loss functions will generally construct different optimal forecasts even though the density for the data is the same for each of them. For example, suppose that the cost of a forecast error, $e = y - f$, is $(1 - \alpha)|y - f|$ for $y < f$ and $\alpha|y - f|$ for $y \geq f$. Then Roger W. Koenker and Gilbert Bassett (1978) show that (in the absence of z) the optimal forecast is $f = F^{-1}(\alpha)$ where F is the cumulative distribution function of a continuous outcome variable Y and F^{-1} is the so-called quantile function. The higher the relative cost of positive forecast errors (higher α), the larger the optimal forecast and hence the smaller the probability of observing costly positive forecast errors.

Two forecasters with different loss functions in this family (different values of α) will want different quantiles of the distribution, so an agency that merely reports a single number could never give them both the optimal forecast. It would be sufficient to provide the entire distribution (density forecast) because this has all the quantile information and hence works for any piecewise linear loss function. Of course, under MSE loss, all forecasters will agree that only the mean of the predictive density is required.

Under the classical approach, a plug-in estimate of θ is generally used to construct the predictive density. The Bayesian equivalent to this approach is to provide the predictive density by removing θ through integration over the prior distribution $\pi(\theta)$ rather than through estimation. The Bayesian chooses $f(z)$ to minimize

$$\begin{aligned}(7)\quad r(\pi,f) &= \int_z \left(\int_\theta \left\{ \int_y \mathcal{L}(f(z),y,z)p_Y(y|z,\theta)\,dy \right\} \right. \\ &\qquad \left. \times p_Z(z|\theta)\pi(\theta)\,d\theta \right) dz \\ &= \int_z m(z) \left(\int_\theta \left\{ \int_y \mathcal{L}(f(z),y,z)p_Y(y|z,\theta)\,dy \right\} \right. \\ &\qquad \left. \times \pi(\theta|z)\,d\theta \right) dz \\ &= \int_z m(z) \left(\int_y \left\{ \mathcal{L}(f(z),y,z) \right. \right. \\ &\qquad \left. \left. \times \int_\theta p_Y(y|z,\theta)\pi(\theta|z)d\theta \right\} dy \right) dz \\ &= \int_z m(z) \left(\int_y \{\mathcal{L}(f(z),y,z)p_Y(y|z)\}\,dy \right) dz.\end{aligned}$$

Now $\int p_Y(y|z,\theta)\pi(\theta|z)\,d\theta = p_Y(y|z)$ is the predictive density obtained by integrating over θ using $\pi(\theta|z)$ as weights. Conditioning on information known to the forecaster, z, the optimal forecast minimizes the bracketed integral.

4. *Loss Functions*

Short of the special (and uninteresting) case with perfect foresight, it will not be possible to find a forecasting method that always sets $f(z,\theta)$ equal to the outcome y. A formal method of trading off potential forecast errors of different signs and magnitudes is therefore required. This is the role of the loss function which describes in relative terms how costly any forecast is given the outcome and possibly other observed data. In mathematical terms, the loss function $\mathcal{L}(f(Z,\theta),Y,Z)$ maps the data, outcome, and forecast to the real number line, i.e., for any set of values of these random variables the loss function returns a single number.

Forecasters thus must pay attention to how errors will affect their results, which means constructing a mathematical representation of potential losses. A natural foundation for a loss function is a utility function that involves both the outcome and the forecast. For a given indirect utility function $U(f(Z,\theta),Y,Z)$, we can set the loss $\mathcal{L}(f,Y,Z) = -U(f,Y,Z)$ in order to see how one might elicit the loss function (Clive W. J. Granger and Mark J. Machina 2006; Spyros Skouras 2001).

4.1 *Determinants of the Loss Function*

Economic insights about the forecasting problem should be used to guide the choice of loss function in any given situation. In particular, the choice of loss should address issues such as (i) the relative cost of over- and underpredicting the outcome variable, i.e., the issue of symmetric versus asymmetric loss; (ii) how economic decisions are influenced by the forecast, which may involve strategic considerations; and (iii) which variables affect the forecaster's loss, i.e., forecast error alone or perhaps the level of the predicted variable matters as well.

4.1.1 *Asymmetric Loss*

On the first point, symmetry versus asymmetric loss, most empirical work in forecasting assumes mean squared error (MSE) loss, which of course implies symmetric loss. Apart from the fact that using MSE loss represents "conventional practice," this choice is likely to reflect difficulties in putting numbers on the relative cost of over- and underpredictions. Construction of a loss function requires a deep understanding of the forecaster's objectives and this may not always be easily accomplished.

Still, the implicit choice of MSE loss by the majority of studies in the forecasting literature seems difficult to justify on economic grounds. As noted by Granger and Paul Newbold (1986, p. 125), ... an assumption of symmetry about the conditional mean ... is likely to be an easy one to accept ... an assumption of symmetry for the cost function is much less acceptable.

Papers that consider the properties of optimal forecasts under asymmetric loss from a theoretical perspective include Granger (1969, 1999), Hal R. Varian (1974), Arnold Zellner (1986), Andrew A. Weiss (1996), Peter F. Christoffersen and Francis X. Diebold (1997), Roy Batchelor and David A. Peel (1998), Granger and M. Hashem Pesaran (2000), Pesaran and Skouras (2002), and Andrew J. Patton and Allan Timmermann (2007a).

Many economic considerations can help in deriving the loss function and determining the extent of any asymmetry. Consider a firm involved in forecasting the sales of a new product. Overpredicting sales leads to inventory and insurance costs and ties up capital. It may also give rise to discounts needed to sell the remaining surplus. Such costs are mostly known or can at least be estimated with a fair degree of precision. Contrast this with the cost of underpredicting sales which leads to stock-out costs, loss of goodwill and reputation, and lost current and future sales. Such costs are less tangible and can be difficult to quantify. Nonetheless, this must be attempted in order to construct forecasts that properly trade off the costs of over- and underpredictions.

For money managers, asymmetric loss may be linked to loss aversion or concerns related to liquidity, bankruptcy, or regulatory constraints (Jose A. Lopez and Christian A. Walter 2001). Under the Basel II accord, banks are required to forecast their Value at Risk, which is a measure of how much they expect to lose with a certain probability, such as 1 percent. Capital provisions are affected by this forecast: overpredicting the Value at Risk ties up more capital than necessary, while underpredicting it could lead to regulatory penalties and the need for increased future capital provisions.

Empirical applications of asymmetric loss include exchange rate forecasting (Takatoshi Ito 1990; Kenneth D. West, Hali J. Edison,

and Dongchul Cho 1993), Budget forecasts (Michael Artis and Massimiliano Marcellino 2001), and the Federal Reserve's Greenbook forecasts (Carlos Carmona Capistran 2006).

4.1.2 *Use of Forecasts*

Turning to the second point, i.e., the use of the forecast, interesting issues in constructing the loss function arise when the forecast is itself best viewed as a signal in a strategic game that explicitly accounts for the forecast provider's incentives. The papers by Tilman Ehrbeck and Robert Waldmann (1996), Marco Ottaviani and Peter Norman Sorensen (2006), David Scharfstein and Jeremy C. Stein (1990), and B. Trueman (1994) suggest more complicated loss functions grounded on game theoretical models. Forecasters are assumed to differ by their ability to forecast. The chief objective of the forecaster is to influence clients' assessment of their ability. Such objectives are common for business analysts or analysts employed by financial services firms such as investment banks or brokerages whose fees are directly linked to clients' assessment of analysts' forecasting ability.

An interesting example comes from financial analysts' earnings forecasts, which are commonly found to be upward biased (e.g., Harrison Hong and Jeffrey D. Kubik 2003 and Terence Lim 2001). By reporting a rosier (i.e., upwards biased) picture of a firm's earnings prospects, analysts may get favored by the firm's management and get access to more precise and timely information. Too strong a bias will compromise the precision of the analysts' forecast and will be detrimental to the position of the analysts in the regular rankings that are important to their career prospects, particularly for buy-side analysts. Ultimately, forecasts must trade off bias against precision. In general, we would not expect the cost of over- and underpredicting earnings to be identical and so biases are likely to persist.

4.1.3 *State Variables*

Turning to the final question, namely which variables should enter into the forecaster's loss function, it is conventional practice to assume that the economic loss only depends on the forecast error, $e = Y - f$. However, this is far too restrictive an assumption in many situations. Patton and Timmermann (2007b) find that it is difficult to understand the Fed's so-called Green Book forecasts of output growth if the loss is restricted to only depend on the forecast error. Rationalizing the Fed's forecasts requires not only that overpredictions of output growth are costlier than underpredictions, but that overpredictions are particularly costly during periods of low economic growth. This finding makes sense if the cost of an overly tight monetary policy is particularly high during periods with low economic growth where it may cause or extend a recession.[3] This is thus an example where $\mathcal{L}(f, Y, Z)$ cannot be reduced to $\mathcal{L}(Y - f)$.

4.2 *Common Loss Functions*

4.2.1 *General Properties*

Loss functions satisfy a set of common properties. Since the perfect forecast, $f(z, \theta) = y$, is the best possible outcome, loss functions achieve a minimum at this point. Thus loss is typically bounded from below at the point where the forecast equals the outcome. In practice, loss functions are usually normalized so $\mathcal{L}(y, y, z) = 0$ for all y and z. For this to be a unique minimum, we have $\mathcal{L}(f(z, \theta), y, z) > 0$ for all $f \neq y$.

Restrictions on the form of the loss function are also needed to make sense of the ideas of minimizing risk, in particular we require that the expected loss exists. The existence of expected loss depends both on the loss function and on the conditional

[3] Central banks commonly state a desire to keep inflation within a band of 0 to 2 percent per annum. Inflation within this band might be regarded as successful outcomes, whereas deflation or inflation above 2 percent are viewed as failures. Again this is indicative of a nonstandard loss function.

distribution of the outcome variable. Recall that expected loss is

$$(8)\quad E_Y[\mathcal{L}(f(z,\theta),Y,z] = \int \mathcal{L}(f(z,\theta),y,z)p_Y(y|z,\theta)\,dy.$$

Hence issues with the existence of expected loss revolve around how large the loss becomes for tail behavior of the predicted variable.

Symmetry of the loss function is the constraint that, for all d,

$$(9)\quad \mathcal{L}(y-d,y,z) = \mathcal{L}(y+d,y,z).$$

Most popular loss functions in economic applications are symmetric.

4.2.2 *Specific Loss Functions*

By far the most commonly employed loss function is MSE loss,

$$(10)\quad \mathcal{L}(f(Z,\theta),Y,Z) = (Y-f(Z,\theta))^2.$$

This is a particularly tractable loss function since there are no unknown parameters and the optimal forecast is simply the conditional mean of Y: $f^*(Z,\theta) = E[Y|Z,\theta]$. Hence under MSE loss the classical "plug in" approach to forecasting simply involves estimating the conditional mean of Y. This relates naturally to regression analysis and the greater part of econometric theory. Under the Bayesian approach, the optimal forecast is the mean of the predictive density, $f(z) = \int yP_Y(y|z)\,dy$.

Mean absolute error (MAE) loss is also very common:

$$(11)\quad \mathcal{L}(f(Z,\theta),Y,Z) = |Y-f(Z,\theta)|.$$

For all continuous distributions $p_Y(y|z,\theta)$, the optimal forecast is the conditional median of Y.

These two loss functions are nested in the general family of loss functions considered by Graham Elliott, Ivana Komunjer, and Timmermann (2005)

$$(12)\quad \mathcal{L}(f(Z,\theta),Y,Z;p,\alpha) \equiv [\alpha+(1-2\alpha)\cdot I(Y-f(Z,\theta)<0)] \times |Y-f(Z,\theta)|^p,$$

where $I(\cdot)$ is the indicator function. For $p = 1$, this gives the lin–lin (piece-wise linear) loss function, which nests MAE loss when $\alpha = 0.5$. For $p = 2$, the asymmetric quadratic loss function, which nests MSE loss when $\alpha = 0.5$, is obtained. Optimal forecasts from the lin–lin loss function are conditional quantiles, while those from the asymmetric quadratic loss function are expectiles (Whitney K. Newey and James L. Powell 1987).

Varian (1974), Zellner (1986), and Christoffersen and Diebold (1997) studied linex loss,

$$(13)\quad \mathcal{L}(f(Z,\theta),Y,Z) = \exp(b(Y-f(Z,\theta))) - b(Y-f(Z,\theta)) - 1,$$

where $b > 0$ is a parameter that controls the degree of asymmetry. If $b > 0$, large underpredictions ($f < y$) are costlier than overpredictions of the same magnitude, with the relative cost increasing as the magnitude of the forecast error rises. Conversely, for $b < 0$, large overpredictions are costlier than equally large underpredictions.

Direction-of-change based loss is another class that has generated considerable interest in recent work. The simplest of these takes the form

$$(14)\quad \mathcal{L}(f(Z,\theta),Y,Z) = \begin{cases} 0 \text{ if } sign(Y) = sign(f) \\ 1 \text{ otherwise} \end{cases}.$$

Sign loss functions are popular in finance since they are closely related to market timing in

financial markets and also are linked to volatility forecasting (Christoffersen and Diebold 2006). To see this, we modify the objective function so that it reflects the magnitude of the outcome variable. For example, consider the decisions of a "market timer" whose utility is linear in the payoff, $U(y, \delta(z)) = \delta y$, where y is the return on the market portfolio (possibly in excess of a risk-free rate) and the action rule, $\delta(z)$, is to invest one unit in the market portfolio if this has a positive expected return and otherwise be short one unit, i.e.,

$$\delta = \begin{cases} 1 & \text{if } f > 0 \\ -1 & \text{if } f \leq 0 \end{cases}. \tag{15}$$

Payoffs from this simple trading rule can be shown to be

$$U(y, \delta(z)) = (2sign(f) - 1)y, \tag{16}$$

and so depend on both the signs of y and f as well as the magnitude of y. Notice that this objective does not adhere to our definition of a loss function. Intuitively this is because large forecast errors for forecasts with the correct sign lead to smaller loss than small forecast errors for forecasts with the wrong sign.[4]

4.3 *Backing Out the Loss Function*

So far we have been discussing how to compute forecasts that minimize expected loss. We next address a different, but related, problem namely how to determine the loss function from a sequence of observed forecasts. Attempts to "reverse engineer" the loss function apply to situations where it is difficult to specify a priori the exact form of the loss function. The idea is to approximate the unknown loss using a flexible family of loss functions such as (12) proposed by Elliott, Komunjer, and Timmermann (2005).

While flexibility is important, parsimony is an equally important concern since it is rare to encounter cases where the number of forecasts amounts to more than a few hundred observations, at least if attention is restricted to the behavior of individual forecasters. Often the situation is even more limited than this. For data sources such as the Survey of Professional Forecasters or the Livingston surveys, individual forecasters with more than a few dozen predictions are fairly uncommon (Elliott, Komunjer, and Timmermann 2006).

Within the context of a given family of loss functions, the unknown parameters of the loss function can be estimated from the forecaster's first order condition (3). For example, for a given value of p, α is the single parameter that controls the degree of asymmetry among the loss functions in (12). When $p = 2$, $\alpha/(1 - \alpha)$ measures the relative cost of positive and negative forecast errors of the same magnitude. For example, a value of $\alpha = 0.4$ suggests that positive errors are two-thirds as costly as negative errors of the same magnitude.

The estimate of α thus provides economic information about the degree of asymmetry required to justify the observed sequence of forecasts. Sometimes such estimates can be rejected on economic grounds. Suppose, for example, that an estimate $\alpha = 0.1$ is required to justify rationality of the observed forecasts. This suggests that it is almost ten times costlier to underpredict than to overpredict the variable of interest. This may be deemed implausible on economic grounds and so asymmetric loss is unlikely to be the explanation for the observed behavior of the forecast error.

How the forecaster maps predictions into actions may also be helpful in explaining properties of observed forecasts. Gordon Leitch and J. Ernest Tanner (1991) studied forecasts of T-bill futures contracts and

[4] In related work, Elliott and Robert P. Lieli (2006) derive the loss function from first principles for binary decision and outcome variables. This setup does not admit commonly applied loss functions for any possible utility function.

found that professional forecasters reported predictions with higher mean squared error than those from simple time-series models. This is puzzling since the time-series models presumably incorporate far less information than the professional forecasts. When measured by their ability to correctly forecast the direction of future interest rate movements—a metric related to the forecasters' ability to make money—the professional forecasts did better than the time-series models. A natural conclusion to draw from this is that the professional forecasters' objectives are poorly approximated by the MSE loss function and are closer to a directional or "sign" loss function. This would make sense if the investor's decision rule is to go long if an asset's payoff is predicted to be positive and otherwise go short.[5]

5. *Estimation of Forecasting Models*

Constructing a forecast for a particular problem requires (i) choosing the variables in z that we intend to employ as inputs in the forecasting model; (ii) taking a stand on the model or set of models for the conditional distribution $p_Y(y|z)$; (iii) specifying how the forecasting model tracks the predicted variable through time, accounting for possible instabilities. The Bayesian approach further requires eliciting priors on the models and their parameters and the classical approach requires an estimator for the unknown elements of θ. The third point is new but will be discussed in more detail in section 6. Here it suffices to say that model instability affects how much weight is put on past as opposed to more recent data in estimating θ.

While economic theory often suggests candidate variables for inclusion in z, rarely is theory so precise as to pinpoint directly which variables should be included or how they should be measured. Economic theory is often better at excluding variables from consideration and limiting the search over which variables to include in the forecasting model.

Moreover, economic theory rarely specifies the distribution or functional form of the model relating y to z. In most forecasting situations, there is therefore considerable uncertainty over the functional form of the model. Finally, the stability through time of the functional form may be questionable. Economies continually evolve, regulations and technology change, suggesting a need to allow the functional form or the parameters of the forecasting model to change over time.

Each of these issues highlights a theme in recent work on economic forecasting which we review below. First, however, we review the workhorse in the forecasting literature, namely the linear forecasting model and its extension to vector autoregressions (VARs). Challenges faced in using VARs for forecasting foreshadow issues that arise for the general problem.

5.1 *The Linear Forecasting Model*

VARs, originally proposed by Sims (1980), constitute a prominent class of forecasting models. Prior to their introduction, larger structural models were the norm in macroeconomic forecasting. While these forecasting models are still employed by central banks and other institutions, the theoretical points of Sims (1980) along with the empirical success of VARs (Robert B. Litterman 1986) has led many forecasters to employ them.

Optimal forecasts are synonymous with linear regression models under MSE loss and a linear specification for the conditional mean. VARs are multivariate extensions of the autoregressive model so commonly used in forecasting and take the form

$$z_t = \beta_0 + \sum_{j=1}^{k} \beta_j z_{t-j} + \varepsilon_t, \tag{17}$$

so $\theta = \{\beta_0, \beta_1, \ldots, \beta_k\}$. In as far as possible, the autoregressive order, k, is chosen so that

[5] Kilian and Simone Manganelli (2006) provide an interesting application to a central banker's forecasts under nonstandard preferences.

ε_t is serially uncorrelated. When $z_t = (y_t, x_t')'$ the first equation of the system is the forecasting equation.

Least squares is the standard plug-in method for estimating the conditional mean from a linear model. Since OLS estimates are consistent and asymptotically efficient for θ when the number of regressors is fixed or grows slowly enough as the sample size increases, they are a reasonable candidate for a plug-in estimator. There are limitations to using this result to justify the use of OLS, however. Because we can write MSE loss as the variance of the forecast error plus the squared bias, the additional loss from using a biased estimator could well be offset by reductions in the variance term. Moreover, since the focus of providing good forecasts is not on the individual estimates for each parameter but a function of these parameters interacted with the data (i.e., $f(z,\theta)$), other estimation approaches could lead to better forecasting performance. Finally, we might not want to rely on asymptotic optimality properties for the OLS estimates of the parameters. In practice, the small sample distributions may differ greatly from their asymptotic counterparts, making comparisons of estimators based on these asymptotic distributions misleading. Taken together, these points have led to a number of other estimation approaches.

Bayesian methods have been suggested as alternatives to OLS in the construction of forecasts from VAR models. Under the additional assumption that ε_t is normally distributed, the likelihood is fully specified and has a well known form. Combined with a set of priors, one can then construct the posteriors and the desired forecasts. For example, the normal prior for the first equation of the VAR results in a closed form solution for the posterior distribution for the parameters β of this equation (In terms of equation (17), β is a vector formed by stacking the columns of $\beta_0, \beta_1 \ldots, \beta_k$). To see this, let $\tilde{Z}$ be a matrix whose rows contain the time series observations of the predictor variables with columns representing the individual regressors and let y be a vector comprising the time series of the predicted variable. In the linear regression model with independently and identically distributed residuals (mean zero and variance $\sigma^2 I$ with σ^2 known) and prior $\beta \sim N(\beta^0, \Omega)$, the posterior distribution for the regression parameters is normal with mean $(\sigma^{-2}\tilde{Z}'\tilde{Z} + \Omega^{-1})^{-1}(\sigma^{-2}\tilde{Z}'y + \Omega^{-1}\beta^0)$ and variance $(\sigma^{-2}\tilde{Z}'\tilde{Z} + \Omega^{-1})^{-1}$. The plug-in forecast simply uses the mean of the posterior, which takes the form of a shrinkage estimator. Setting $\Omega = \sigma^2 k^{-1} I$, the estimator becomes $(\tilde{Z}'\tilde{Z} + kI)^{-1}\tilde{Z}'y$, which is in the form of a Ridge estimator. Under MSE loss, normally distributed $\hat{\beta}$ and any general prior distribution $\pi(\beta)$, it can be shown that the Bayes rule forecast in a linear regression takes the form of a correction to the OLS estimator.

To employ these methods in practice requires specifying the prior (and typically the results are extended beyond the known variance case). Litterman (1986) and Thomas Doan, Litterman, and Sims (1984) suggested Minnesota priors on the parameters of a VAR which more heavily weight the parameter configuration toward a model where variables follow individual random walks. More distant lags are shrunk toward zero more heavily. This type of prior can be helpful in obtaining better forecasts of macroeconomic outcomes. John C. Robertson and Ellis W. Tallman (1999) examine the forecasting performance of flat prior VARs (i.e., the usual OLS estimates) versus Bayesian methods based on more informative priors. They find that extensions to the Litterman priors revolving around long run properties provides forecasting gains for a number of macroeconomic variables (GDP growth, unemployment, Fed funds rate, and CPI inflation). K. Rao Kadiyala and Sune Karlsson (1993, 1997) examine more extensive priors than the Litterman approach (allowing for dependence between the equations) and find examples of improvements over the Minnesota prior.

A promising recent literature uses DSGE models to constrain VARs in a Bayesian setting. In this vein, Marco Del Negro et al. (2007) cast DSGE models as (reduced-form) VARs that include an error correction term. Theoretical restrictions from firms' and households' optimizing behavior subject to their intertemporal budget constraints, along with assumptions about government expenditures imply a set of cross-equation restrictions on the parameters of the VAR. Del Negro et al. (2007) relate these constraints to the priors on the model parameters; ignoring the theory corresponds to diffuse priors while informative priors pull the parameters towards the theoretical constraints. In a simulation study, these authors find evidence that, for a range of macroeconomic variables, the DSGE-based VAR produces better out-of-sample forecasting performance than the standard unconstrained VAR.

5.1.1 *Shrinkage Methods*

A variety of estimation and variable selection methods have been suggested as plug-in estimators to attain a better trade-off between the bias and variance components of the forecast MSE. Most of the alternative plug-in estimators are modifications of the OLS estimator, $\hat{\beta}_i^{OLS}$, and fall in the general class of shrinkage estimators, $\hat{\beta}_i^s$, of the form

$$\hat{\beta}_i^s = (1 - \gamma_i)\hat{\beta}_i^{OLS} + \gamma_i\tilde{\beta}_i \qquad (18)$$

$$0 < \gamma_i < 1, i = 1,\ldots,k,$$

where $\tilde{\beta}_i$ is the shrinkage target. The shrinkage weight, γ_i, generally depends on the data (W. James and Charles Stein 1961). It is common practice to set $\tilde{\beta}_i = 0$, in which case the OLS estimators are shrunk toward zero. The expected gain from this approach under MSE is to reduce the variance in the bias-variance trade-off that arises from the plug-in estimator for the regression coefficients. For empirical evidence on shrinkage methods, see Zellner and Chansik Hong (1989).

Estimators differ in how they specify the shrinkage weights γ_i and shrinkage target $\tilde{\beta}_i$. These include bagging and subset selection methods (explained in detail in section 5.2.1 below), as well as Stein regression and many Bayes and empirical Bayes methods. Other examples of shrinkage methods that have been less employed for forecasting include ridge regression (which is a special case of (18)), as well as the lasso (Robert Tibshirani 1996) and the (non negative) garrote (Leo Breiman 1995). The latter two methods are somewhat more complicated in estimation and are essentially penalized least squares estimation methods. For the lasso the data are first normalized (see Alan Miller 2002 for a textbook treatment) and then the sum of squared errors from the regression is minimized subject to the constraint that the sum of the absolute value of the regression coefficients is less than a chosen value. For the garrote, first least squares estimates from a regression of y_{t+1} on x_t are obtained, then values c_i are chosen to minimize

$$\sum_{t=1}^{T-1}\left(y_{t+1} - \sum_i c_i\hat{\beta}_i^{OLS}x_{it}\right)^2,$$

subject to the constraints that each $c_i \geq 0$ and that their sum falls below a chosen value.

5.1.2 *Estimation with Non-Standard Loss Functions*

It is common to use standard least-squares estimation techniques for model parameters without reference to the loss function. An example is using linear regression and examining if the forecasts capture turning points in the data. Despite such practice, recent studies have suggested that forecasting performance can often be significantly improved by using the same loss function in the estimation and evaluation stages. This has been found both in simulation experiments (Elliott and Timmermann 2004) and in empirical studies (Christoffersen and Kris Jacobs 2004).

To understand such gains, note that, for loss functions other than MSE, optimal forecasts generally require use of the loss function either directly in estimation or through the specification of the model that is estimated. For example, in the case of lin–lin loss, the optimal forecast is a conditional quantile and so quantile methods can be employed to estimate the parameters of the forecasting model. When this is specified up to a set of unknown parameters—i.e., the function $f(.)$ of the forecasting model, $f(z,\theta)$, is known but θ is not—then the loss function can be used to estimate the unknown parameters through M estimation, i.e.,

$$(19)\quad \hat{\theta} = \arg\min_{\theta} (T-1)^{-1} \sum_{t=1}^{T-1} \mathcal{L}(f(z_t,\theta), y_{t+1}, z_t).$$

When the loss function is differentiable, extremum estimators based on the first order moment conditions can also be used. Approximate methods have been suggested by Weiss (1996). In the context of forecast combination, Elliott and Timmermann (2004) propose methods of moments estimators and approximate methods for a variety of loss functions including lin–lin and asymmetric quadratic loss.

5.2 *Estimation with Many Variables*

The inability of economic theory to precisely define which variables should be included in a forecasting model has become an important issue since thousands of potentially relevant variables are readily available from governments and other organizations. The virtual explosion in the number of potential predictor variables is exacerbated by the fact that the dynamic structure of the forecasting model is typically unknown. Adding an extra variable therefore increases the model dimension not only by a single parameter but by many parameters to account for the dynamic effects of this variable on the outcome. At the same time, the length of the available data is often relatively short because the frequency of time series observations and the period over which data have been constructed combine to limit the sample size. Model instability (which will be discussed below) may further limit the useful length of the data.

Short samples, whether due to limited data availability or model instabilities, along with large sets of predictor variables mean that forecasters always face a trade-off in terms of how complicated the model can be versus how well its parameters can be estimated. Methods that place constraints on the number of parameters may produce better forecasts even if the population model for the data does not share these constraints. For example, small and very parsimonious models have been found to work well in many empirical studies. Reductions in loss resulting from adding more variables are often more than offset by the resulting increase in parameter estimation error. For MSE loss, additional variables reduce bias, but estimation error increases the variance, leading to a bias-variance trade-off. Methods other than omitting relevant variables and using OLS are available to exploit this trade-off. This has led to the proposal of a wide range of estimation techniques in the forecasting literature that we next turn to.

5.2.1 *Subset Selection Methods*

Different forecasting methods amount to different weighting functions on the underlying regressors (Stock and Watson 2005). Methods that put full weight on all regressors tend to suffer from imprecisely estimated parameters. At the opposite extreme are purely autoregressive methods that ignore information in other variables and focus solely on a variable's own history.

A natural approach in the middle of these extremes is to only use a subset of the available predictor variables for forecasting. Inclusion of additional variables in the regression model increases the variance of the forecast, although if their true coefficients are nonzero then this also reduces the bias. Removing variables with coefficients that are small enough (so the resulting bias is small)

to avoid this additional increase in variance may yield better performing forecasts. Of course, if the coefficient on the omitted variable were truly zero, then there is no bias and the variable should always be excluded.

Subset regression methods set $\tilde{\beta}_i = 0$ in (18), while γ_i becomes an indicator function based on the adopted rule for variable inclusion or exclusion. Various methods for choosing the regressor subset are in use. When there is a natural ordering to the regressors (e.g., in vector autoregressions) penalized likelihood methods such as the Akaike information criterion (AIC) or the Bayesian information criterion (BIC) have been employed. Except for cases with very small sample sizes the BIC applies a heavier penalty term than the AIC and so this method includes a smaller subset of regressors than AIC. Such selection methods are less common without a natural ordering because the set of models to search over easily becomes very large without this ordering. For example, with 30 potential variables, there are 2^{30} or more than one billion different models that have to be considered.

Atsushi Inoue and Kilian (2006) discuss conditions under which a variety of tools for model selection—e.g., ranking by recursive MSE, rolling MSE, or model selection by the AIC or BIC—will identify the model with the lowest true out of sample MSE among a finite set of forecasting models. They find that selection by AIC and ranking by recursive MSE yield inconsistent results and have a positive probability of choosing a model that does not have the best forecasting performance while the SIC is consistent for nested models. Of course, it should be borne in mind that consistency is not the most important criterion to satisfy here and how a method handles bias–variance tradeoffs may be more important to forecasting performance.

An alternative approach is to evaluate each variable on its own or in smaller groups—a method advocated by, e.g., Hendry and Hans-Martin Krolzig (2004). This step-wise procedure employs t-tests to remove individual variables with statistically insignificant coefficient estimates. While computationally highly attractive, this approach gives rise to problems of its own: insignificant estimates can arise not only because the true parameters are small but also because of large sampling error. Moreover, if one reestimates the model after removal of some parameters and again examines statistical significance, the method becomes path dependent. This matters because classical pretests need not result in consistent estimates of the model. Finally, the "all or nothing" approach of either using the OLS estimate or omitting the regressor may be too restrictive (i.e., restricting γ_i at zero or one).

Methods that attempt to exploit the bias-variance trade-off without the "all or nothing" approach have thus been suggested. The estimation method closest to pretesting is bagging (bootstrap averaging) proposed by Breiman (1996). In this method, the forecast model is bootstrapped, reestimated using tests for significance to omit variables, and the forecast from the bootstrapped model is computed. A final forecast is then computed as the average of the forecast over the bootstrapped models (see Inoue and Kilian (forthcoming) for a more complete description of the method). Since variables are unlikely to be omitted for all bootstrap replications, the estimator can be viewed as a smoothed version of the "all or nothing" approach. Bagging still sets $\tilde{\beta}_i = 0$ but γ_i is now equal to the average over this indicator function across the bootstrap replications. In this way bagging changes from a hard threshold (zero or one) to a soft one (some number between these values).

5.2.2 *Factor Models*

An alternative to excluding variables from a large dimensional set of predictors is to extract common features from the data and then use these as the basis for the forecasting model. Indeed, the dominant classical approach for dealing with large dimensional

data is to extract a set of common factors of much lower dimensionality than the original variables to summarize an otherwise overwhelming amount of information. Suppose Z contains N economic variables whose common dynamics can be represented through the factors

$$Z_t = \Lambda(L)\psi_t + e_t, \tag{20}$$

where e_t is a vector of idiosyncratic shocks, ψ_t is a vector of common factors, and $\Lambda(L)$ is a matrix of lag polynomials representing dynamic effects. For low dimensional systems (small N), dynamic factor models can be estimated through the Kalman filter. When N is large, Stock and Watson (2002) propose a principal components approach to obtain the common factors as the solution to a simple least squares problem. An alternative approach proposed by Mario Forni et al. (2000, 2002) is to extract principal components from the frequency domain using spectral methods.

While the construction of a set of common factors resolves the question of how to aggregate an otherwise far too large dimensional state vector, use of these techniques in forecasting also raises new issues. There is a risk that the factor extraction serves as a "black box" approach void of any economic interpretation. This risk arises in situations where the factors are not clearly identifiable with underlying blocks of economic variables, although in many situations such blocks can be used to good avail for interpretation purposes (Sydney C. Ludvigson and Serena Ng 2005, 2007). The aim is to ensure that the first few factors can be interpreted in a way that links them to a particular subset of variables.

In an empirical analysis, Stock and Watson (2005) find that methods that include the first few (and most significant) principal components in addition to own-variable autoregressive dynamics generally work best and that a few principal components are responsible for most of the improvement in forecasting performance.

5.3 *Choice of Functional Form: Nonlinear Models*

There is little reason to expect that the economy yields linear relationships between the data and the predicted variable. Indeed, empirical tests for various forms of nonlinearity often reject linear benchmark models (Timo Teräsvirta 2006). For example, important nonlinearities have been identified in the behavior of asset returns, particularly in the large literature on volatility modeling and forecasting. Hence it makes sense to explore (if not necessarily employ) nonlinear models as a way to improve the forecasting performance of linear models. Typically the literature on nonlinear forecasting assumes MSE loss so the focus remains on estimating or approximating the conditional mean of the predicted variable.

Most nonlinear models in common use take the form

$$y_{t+1} = \theta_1' z_t + g(z_t, \theta_2) + \varepsilon_{t+1}. \tag{21}$$

Assuming that $E[\varepsilon_{t+1}|z_t] = 0$, this nests the linear model when $g(.) = 0$.

Extending the set of forecasting models under consideration to nonlinear specifications substantially expands the model set. Let the set of models, $\mathcal{M}$, represent the combinations of parameters and functional forms for the models under consideration. When the true model $M_0 \in \mathcal{M}$, the model is said to be correctly specified, otherwise it is misspecified. A misspecified model may yield forecasts that are difficult to beat, even if the coefficients are not meaningful (Clements and Hendry 1998). For example, a linear forecasting model estimated by OLS results in the linear model that minimizes Kullback Leibler distance between the estimated "approximate" model and the unknown nonlinear model. That said, however, when forecasters have to "sell" their forecasts to decisionmakers, the misspecified model may be difficult to put a story to.

The forecaster's problem is to choose the best available model $M \in \mathcal{M}$. Because this

involves a search over functional forms, to be able to actually perform this search the forecaster needs to restrict the problem further. As with the specification of the variables to be included in the regression model, economic theory is often not particularly precise on the exact form of the nonlinearity to be expected.

Tests for the functional form $g(.)$ are complicated because θ_2 or a subset of this vector is not identified under the null hypothesis of linearity. As a result, standard methods such as the generalized likelihood ratio test lose optimality properties and are no longer approximated asymptotically by a chi-squared distribution. A large literature has arisen to deal with these issues (see Teräsvirta 2006).

The danger of using a misspecified forecasting model is a real problem given the lack of theoretical underpinning of the choice of $\mathcal{M}$. Moreover, often $\mathcal{M}$ is a very small subset of the possible model specifications. Furthermore, for many of the tests of the null of linearity for these models, rejection does not necessarily imply that a particular nonlinear model chosen is implied. Hence it is important to understand the estimation and forecasting properties of these models under misspecification. Such misspecification and the approximation properties of nonlinear models may explain why they sometimes generate extreme forecasts.

The literature has pursued two broad themes: (i) the formalization of specific nonlinear models that generalize the linear model in an intuitive way; or (ii) the use of more global approximation procedures that seek to approximate the unknown nonlinear function. Which approach should be adopted depends on how much is known about the type of nonlinearity to be expected in a given situation. We next examine each approach in turn.

5.3.1 *Local Approximations*

One branch of the literature considers nonlinearities that are essentially ad hoc models designed to be more flexible than linear models which arise as special cases. The set of models $\mathcal{M}$ is typically relatively small, nesting a linear specification, and fully defined up to a set of unknown parameters. Much of this work has been to extend autoregressive models and has its roots in the historical domination of autoregressive integrated moving average forecasting models (George E. P. Box and Gwilym M. Jenkins 1970) of the form

$$\phi(L)\Delta y_{t+1} = \eta(L)\varepsilon_{t+1}. \tag{22}$$

Here $\phi(L)$ and $\eta(L)$ are lag polynomials while $\Delta y_{t+1} = y_{t+1} - y_t$ takes first differences and ε_{t+1} are serially uncorrelated innovations (white noise). Examples include the family of smooth transition autoregressive (STAR) models, which typically set $z_t = (y_t, y_{t-1}, \ldots, y_{t-p})$ and differ in the exact form for $g(.)$ as specified in (21), e.g. the exponential STAR model sets elements of $g(.)$ equal to $\theta_{21}'z_t(1 - \exp(-\theta_{22}'z_t))$ whereas the logistic STAR (LSTAR) model sets $g(.)$ equal to $\theta_{21}'z_t(1 + \exp(-\theta_{22}'z_t))^{-1}$, commonly with restrictions on θ_{21}, θ_{22}. Similar nonlinear models can be used with specifications for z_t other than lagged dependent terms in situations where a candidate variable explaining the nature of the nonlinearity (e.g., financial crises as measured by default premia) is available.

Threshold and regime switching regressions set $g(.)$ equal to $\theta_{21}'z_tI(s_t < \theta_{22})$ for some state variable, s_t, where $I(.)$ is the indicator function. Smooth threshold models have been employed empirically with some success to forecast real exchange rates (Lucio Sarno and Mark P. Taylor 2002), industrial production and a host of other macroeconomic series (Teräsvirta, Dick Van Dijk, and Marcelo C. Medeiros 2005).

When the state is unobserved, it is common to model it through a regime switching process (James D. Hamilton 1989). The literature predominantly uses fully parameterized Gaussian mixture models with weights on the individual states that are determined

from the updated state probabilities. We cover these models in more detail in section 6.2 on breaks.

The large literature on autoregressive conditional heteroskedasticity (ARCH) in asset returns (reviewed in the context of forecasting by Torben G. Andersen et al. 2006) is another example of nonlinear dynamics that could be important to portfolio managers concerned with predicting returns and managing the risk of their assets. These models imply that the conditional variance of asset returns is persistent and hence partially predictable, particularly at short horizons. Such predictability of the conditional variance can be captured by extending (21) to

$$E[\varepsilon_{t+1}^2|z_t] = h(\varepsilon_t, \varepsilon_{t-1}, \ldots, \varepsilon_1).$$

Nonlinear forecasting models can adapt more quickly to changes in the underlying time-series dynamics and avoid smoothing the data as much as linear models. This same feature means that nonlinear models can also be highly sensitive to the sort of outliers found in many economic and financial time series and may be more prone to overfitting than linear models. Furthermore, the parameters capturing nonlinear dynamics are often associated with a few episodes such as the change in the dynamics of U.S. interest rates during the "monetarist experiment" from 1979 to 1982. As a consequence, these parameters can be very imprecisely estimated for the typical sample sizes available to macroeconomic forecasters and so these models often produce quite poor out-of-sample MSE performance.[6] On the other hand, nonlinear forecasting models may perform quite well in certain states (e.g., recessions or periods of financial crises) and so can be used either in conjunction with other models that generate more smooth forecast (see section 8.4 on combination) or for nonconvex loss functions such as the sign function (14) that put smaller weight on outliers.

[6] Indeed, some simulation studies find that even when the nonlinear model is correctly specified, it often produces less precise forecasts than a simple misspecified linear approximation due to the greater uncertainty about the nonlinear model's parameters (Zacharias Psaradakis and Fabio Spagnolo 2005).

5.3.2 *Flexible Approximations*

Another approach to forecasting with nonlinear models has been to acknowledge the uncertainty over the functional form, $g(.)$, and attempt to construct a flexible approximation by considering a very large set of models to include in $\mathcal{M}$. To make the search over models operational, rather than search over all of $\mathcal{M}$, this approach searches over an approximating set $\tilde{\mathcal{M}}$ of the form

$$(23) \quad y_{t+1} = \theta_1' z_t + \sum_{j=2}^{J} g_j(z_t, \theta_j) + \varepsilon_{t+1}.$$

Often the simplification $g_j(z_t, \theta_j) = \tilde{g}_j(z_t)\theta_j$ is employed to make the forecasting model linear in the parameters (or at least more so, since additional parameters can be hidden inside $\tilde{g}(.)$). The idea is to choose the functions $g(.)$ carefully enough and the number of them, J, large enough to approximate a wide variety of possible nonlinear functions, see e.g., Norman R. Swanson and Halbert White (1995).

There are a large number of theoretically well motivated choices for the basis functions $g_j(z_t, \theta_j)$. Most popular in the economic forecasting literature are artificial neural network models, where $g_j(z_t, \theta_j) = \theta_{j1}(1 + \exp(-z_t'\theta_{j2}))^{-1}$ and various methods are employed for choosing or estimating θ_{j2}. Other basis functions include Fourier series, polynomials, piecewise polynomials and splines. Methods such as wavelets, ridgelets, and the Gallant (1981) flexible fourier form also belong to this set. Both theoretically and in practice, different methods work well against different classes of functions.

Practical problems arise for these methods in terms of estimation and forecasting. First, unless $\tilde{g}(.)$ is fully specified, estimation requires nonlinear optimization. Variations

have arisen in attempts to find simple methods to specify the functional form, so as to leave the remaining estimation linear in the parameters and hence estimable by OLS. This is standard for example in the application of neural net models.

Second, the order J must be chosen. Since an infinite number of possible terms could be included and the in-sample fit is improved by choosing J as large as possible, overfitting is likely unless the number of included terms is somehow restricted. Overfitted models tend to produce very good results for the data used to estimate the models, but very poor forecasts on fresh (out-of-sample) data. To deal with these issues, methods such as information criteria and cross validation are used to select among this class of models. Overfitting remains the Achilles heel of these methods, however.

Finally, as with parametric nonlinear models, the risk of generating relatively extreme forecasts remains when forecasting from sample points where the data is relatively sparse. Many practitioners use "insanity" filters, replacing these forecasts with a smoothed value when the forecast is too far from the outcome or its mean. Even so, the track record of these models in forecasting has been mixed. Nonlinearities do seem to be present in many macroeconomic series, but the data samples for these variables tend to be relatively short, thus hampering the precise estimation of nonlinear forecasting models.

For financial returns, the signal-to-noise ratio—i.e., the fraction of predictable variation in asset returns—tends to be very low. Often this means that the observed nonlinearities are poorly identified, imprecisely estimated and so the risk of overfitting is very high. As a result, the deterioration in out-of-sample forecasting performance is likely to be very high when compared against the predictive performance during the training sample used to estimate the parameters of the model (see, e.g., Jeffrey Racine 2001). Consequently there is little evidence that such forecasting models dominate simple linear specifications, at least under MSE loss.

Overall, the difficulties that arise when forecasting with nonlinear models revolve around the question whether a fitted nonlinear model provides a good approximation to the true nonlinear model. In the case of the ad-hoc specifications, rejection of a linear model in favor of a particular nonlinear specification does not necessarily indicate that the latter will produce good forecasts. Rejections of typical tests for nonlinearity tend to indicate a range of possible models rather than a particular model. In the case of model approximation, model specification search tends to result in models that overfit the data, again causing problems for forecasting.

5.4 *Multiperiod Forecasts*

In many forecasting situations, the forecast horizon, h, is longer than the frequency used in collecting data and estimating the forecasting model. For example, a central bank may be interested not only in forecasting short-run inflation but also in inflation over the medium and long run. In these situations, the forecaster encounters a multiperiod forecasting problem.

Computing multiperiod forecasts is simple if the predictors are weakly exogenous but issues arise in practice. To illustrate this point, when forecasting from a univariate first-order autoregressive model $y_t = \phi y_{t-1} + \varepsilon_t$, forward iteration gives

$$y_{t+h} = \phi^h y_t + \sum_{i=0}^{h-1} \phi^i \varepsilon_{t+h-i}. \tag{24}$$

Assuming that $E[\varepsilon_{t+j}|y_t] = 0$ for $j > 1$ and that both the model and its parameters are known, the optimal forecast under MSE loss is simply $\phi^h y_t$.

When the parameters are unknown, however, the problem becomes far more complicated. A simple solution would be to use the plug-in OLS estimate, $\hat{\phi}^h y_t$. However, this is clearly only a solution of convenience: $\hat{\phi}$ is

generally not unbiased and thus, $\hat{\phi}^h$ will not be unbiased for ϕ^h either. Even if $\hat{\phi}$ were unbiased, in general $\hat{\phi}^h$ would not inherit this property.

An obvious alternative to iterating forward on a single-period model is to tailor the forecasting model directly to the forecast horizon. This is more in spirit with viewing forecasting models as misspecified simplifications of the underlying data generating process and entails a model of the form

$$y_{t+h} = g(z_t, \theta) + \varepsilon_{t+h}. \tag{25}$$

The chief problem is now the overlap in the forecast errors that will generally exhibit behavior similar to that of a moving average process of order $h - 1$. For example, if $h = 2$, the forecast errors will be serially correlated according to an MA(1) process even if the true forecasting model is used and its parameters are known. Such serial dependence can be handled through a number of procedures that account for autocorrelation in the forecast errors.

Which approach is best—the direct or the iterated—is an empirical matter since it involves trading off estimation efficiency against robustness to model misspecification. It is also not clear how iterating multiple periods ahead on a misspecified model will affect the quality of the forecast. For example, the initial value of the conditioning information (z_t) could well matter in this situation. Even when the models are correctly specified, there is a trade-off between the cumulative effect on the forecast of using plug-in parameter estimates (which is avoided in the direct approach), versus the greater efficiency of the iterated approach that comes from estimating the forecasting model on data measured at a higher frequency.

Marcellino, Stock, and Watson (2006) address these points empirically using a data set of 170 U.S. monthly macroeconomic time series. They find that the iterated approach generates the lowest MSE values, particularly if long lags of the variables are included in the forecasting models and if the forecast horizon is long. This suggests that reducing parameter estimation error can be more important than concerns related to model misspecification, an issue that cannot be decided ex ante on theoretical grounds alone (Frank Schorfheide 2005).

Special problems may arise when forecasting multiple steps ahead with nonlinear models, where numerical methods are typically required due to the nonlinear form. This problem stems directly from taking the ad hoc model to be the true model, which is or course a doubtful assumption. To illustrate this, suppose that

$$y_{t+1} = g(y_t; \theta) + \varepsilon_{t+1}. \tag{26}$$

Iterating forward to the two-period horizon, we have

$$y_{t+2} = g(g(y_t; \theta) + \varepsilon_{t+1}) + \varepsilon_{t+2}. \tag{27}$$

Hence the function g (presumed to be nonlinear) needs to be invoked as many times as the length of the forecast horizon. Moreover, the entire distribution of ε becomes crucial even under MSE loss where interest is limited to forecasting the conditional mean. For example, if g is quadratic, the variance of ε matters to forecasting the conditional mean two or more periods ahead.

6. *Model Instability*

Economic institutions, tax rules, and political regimes change over time and the economy evolves in response to technological and macroeconomic shocks such as the oil price changes in the 1970s. One of the stylized facts of empirical macroeconomics is the "Great Moderation," i.e., the lower volatility of many macroeconomic series after the mid-1980s. Events such as these make it plausible that the underlying data generating process changes over time. In many forecasting situations, economic theory is silent on the exact form of these instabilities and instead offers

general guidelines that can help determining the source (and possibly timing) of instabilities such as changes in monetary policy (due, e.g., to a change in the Federal Reserve chairman) or changes in economic institutions.

In the construction of the forecasting problem in section 3, the specification of the likelihood for the data $p_Y(y|z)$ did not require that the relationship between the data remains stable over time or that the underlying data itself is stable through time, although a model for the process that generates instability is required. A difficulty that arises in the presence of changes in the data generating process is the existence of a multitude of models that can capture potential instabilities. Unit root models are popular, although the root could be near one rather than exactly equal to one. Fractionally integrated models allow similar behavior at low frequencies. Breaks in regression parameters can also mimic this type of behavior. Beyond this we could allow the root to be stochastic and near one.

Model instability introduces at least three problems for forecasters. First, it complicates specification of the likelihood for the data. From a Bayesian perspective, this can make it more difficult (at least analytically) to determine a closed form forecasting rule, depending on the form of the nonstationarity. Second, since the parameterization of the nonstationarity results in a larger dimension of θ, estimation is also affected. Finally, nonstationary data makes averaging over the past to obtain plug-in estimates more difficult. In classical estimation, this can be a large problem. Further complications arise through nonstandard properties of the estimators that frequently arise in these models.

6.1 *Breaks in Model Parameters*

The existence of nonconstant parameters in many of the data series that economists have an interest in forecasting is well documented. Stock and Watson (1996) considered the stability of a large set of macroeconomic and financial variables and found evidence of model instability for the majority of these. Clements and Hendry (2006) also stress instability as a key determinant of forecasting performance.

Most work on forecasting models with unstable parameters has considered linear specifications of the form

$$Y_{t+1} = (\beta_t - \bar{\beta})Z_{1t} + \gamma Z_{2t} + u_{t+1}, \tag{28}$$

where the coefficients β_t on Z_{1t} are changing over time while the remaining coefficients are constant.

The first problem that arises in the construction of a forecasting model of this type is that there are many ways in which β_t can be nonconstant. We could parameterize β_t as a stochastic process (either mean reverting or not) or as a step function that changes at random times by random amounts. Examples of such models include the popular unobserved components model (where $Y_{t+1} = \beta_t + u_{t+1}$ and β_t follows a random walk process) and extensions of this to the entire vector β_t (as in the models of Mike West and Jeff Harrison 1997).

If the variation in β_t is not permanent in the sense that it can be characterized by a mean reverting process, the linear specification that omits the breaks in β_t is essentially a heteroskedastic model and least squares estimation of the parameters will not be too misleading (White 2001).

When the breaks are permanent, the coefficients of the linear model lose meaning and become similar to sample averages of a random walk, changing with time and not related directly to any parameter of the model. In either case, knowing the true model will enable better forecasts.

Forecasters must decide whether or not to (a) use only part of the data available, assuming that the retained data is sufficiently stationary that it will provide a good approximation to a model with constant coefficients, or (b) attempt to model the breaking process.

To illustrate these approaches, suppose β_t is constant apart from a single break of unknown size δ at an unknown date τ,

$$(29) \quad Y_{t+1} = \begin{cases} \beta Z_{1t} + \gamma Z_{2t} + u_{t+1} & t < \tau \\ (\beta + \delta) Z_{1t} + \gamma Z_{2t} + u_{t+1} & t \geq \tau \end{cases}.$$

An example of the first approach would be to try and estimate $\hat{\tau}$ and base the estimates of the forecasting model solely on data after the break. Alternatively, a forecaster might consider constructing estimates for both the break date $\hat{\tau}$ and the size of the break $\hat{\delta}$ in order to construct a forecast from the full data incorporating the break into the model (and possibly attempt to forecast future breaks). This would be an example of the second approach.

Unfortunately, while tests for nonconstant parameters are quite good at detecting breaking behavior of this nature, they are not capable of distinguishing the particular type of nonstationarity beyond the distinction of "permanent" deviations versus the mean reverting deviations mentioned above. Nearly all popular tests have no power against temporary deviations of β_t from its mean. Conversely, nearly all tests have similar power against a host of possible processes for β_t when it does depart permanently from any value. This is true for models with few breaks, many breaks, or breaks every period.[7]

The implication for forecasting is that once one has found evidence of breaks in the parameters of some variables, there is still a great deal of uncertainty as to the nature of the breaking process. Because it will be difficult to pin down the appropriate model, parameterizing and estimating the breaking process will generally be quite difficult.

[7] Stock and Watson (1998) show that tests for a single break have power against random walk breaks. Elliott and Ulrich K. Muller (2006) consider a wide class of breaking processes and show that optimal tests for each of the breaking processes have equivalent asymptotic power against all of the other breaking processes in a wide class.

Even if it were known that the forecasting model has a single break point, estimates of the break date are often not particularly useful in practice. Tests for a break will often reject stability even though the break size is too small to permit precise estimation of exactly when the break occurred. One can still proceed with the first approach and try to estimate the window of data to use for the forecast. However, some account for the uncertainty surrounding the timing of the break is likely to be an important part of a successful forecasting strategy in the presence of breaks.

For the alternative of estimating the full model, including both the break date and the size of the break, Elliott (2005) shows that estimation of the break size and its location results in very poor forecasts relative to knowing these parameters. Instead, a method of averaging over all possible break dates with weights that depend on sample estimates of the probability that each date is the true break date is suggested, with substantial gains over least squares estimates of these two parameters. In an empirical application to exchange rate forecasting, Barbara Rossi (2006) found evidence of widespread instabilities and showed that Elliott's (2005) method works well in practice for forecasting a range of currencies.

Similarly, Pesaran and Timmermann (2005b, 2007) find that the estimation window matters significantly to the out-of-sample forecasting performance of simple time-series models in the presence of breaks. Given the considerable uncertainty surrounding the time and the size of the break, they consider approaches that average forecasts generated under different estimation windows. They also derive analytical results for the normal model under MSE loss and demonstrate that the gains in forecast accuracy from using pre-break data increases when breaks are small and occur late in the sample.

When there is more than one break, things become even more difficult. Jushan Bai (1997) and Bai and Pierre Perron (1998) suggest an

approach to determine the number of breaks through repeated tests on the data. This approach has been applied to forecast stock returns by Bradley S. Paye and Timmermann (2006) and David E. Rapach and Mark E. Wohar (2006). Their results suggest the presence of multiple breaks in standard forecasting models for stock returns and reveal wide variation in the extent of predictability in stock returns across break segments. Paye and Timmermann (2006) also find that the break dates are difficult to pin down, vary greatly across different model specifications, and do not seem to be common across international markets. This makes the task of forecasting stock returns particularly difficult since the question of "how much historical data to use" and how to weight new versus old data is both very important in practice and difficult to come up with a satisfactory solution to.

Structural breaks in parameters can also cause a forecasting model's performance to deviate significantly and erratically from the outcome expected on the basis of its in-sample fit. Rossi and Raffaella Giacomini (2006) refer to these situations as model breakdowns and develop a method to empirically detect them.

The two most common types of instability found in macroeconomic and financial data are breaks in the model parameters and unit root or long memory behavior of the data. We next discuss each of these.

6.2 *Modeling the Break Process*

The presence of historical breaks in a time-series model requires that the possibility of future breaks be considered. This means that the process generating breaks must itself be modeled. In this regard, the forecasting problem is unique compared with the problem of detecting and dating past breaks. Approaches that do not model the break process itself and treat breaks as deterministic (such as Bai and Perron 1998) are not directly applicable to forecasting.

This is not a problem for the time-varying parameter specifications which directly posit a model for how the parameters evolve in future periods. A popular approach is to parameterize β_t as a random walk and use the Kalman Filter to estimate the path for β_t and produce a forecast (Andrew Harvey 2006 covers the classical approach while West and Harrison 1997 cover the Bayesian approach).

The simplest example arises when $Z_t = 1$, β_t is a random walk and both the innovations to β_t and Y_t are normally distributed, so the forecast of Y_{T+1} is $\hat{\beta}_T$. Then

$$\hat{\beta}_t = \hat{\beta}_{t-1} + \phi_t(Y_t - \hat{\beta}_{t-1}), \tag{30}$$

where ϕ_t depends on the variances of the two error terms. For a given choice of initial value, this recursion can be used to generate forecasts in real time. In the limit, ϕ_t can be approximated by a constant, which for a given value of ϕ yields the exponentially weighted average (IMA(1,1)) model

$$\hat{\beta}_t = \frac{(1-\phi)}{(1-\phi^t)} \sum_{s=0}^{t-1} \phi^s Y_{t-s}. \tag{31}$$

This approach is equivalent to the discounted least squares model which puts a decreasing weight on data further back in time. These methods can readily be extended to the general model with time-varying predictor variables.

Examples of empirical application of these models are plentiful and include tracking of the skills of mutual fund managers (Harry Mamaysky, Matthew Spiegel, and Hong Zhang 2007) and prediction of variables, such as GDP growth, inflation, and electricity demand (Harvey and Siem Jan Koopman 1993).[8]

Another example is the recurring stochastic breaks models embodied in the Markov switching approach of Hamilton (1989). This assumes that changes to the parameters of the

[8] Risk Metrics use this method to track conditional volatility in financial markets and typically sets ϕ close to one.

model are driven by a latent state variable, S, that follows a first-order Markov chain. For example, the linear forecasting model could be modified as follows:

$$y_{t+1} = \beta_{s_{t+1}} z_t + \sigma_{s_{t+1}} \varepsilon_{t+1}, \tag{32}$$

where the dynamics in the state variable takes the form

$$\Pr(S_{t+1} = j | S_t = i) = p_{ij}, \quad i,j = 1,\ldots,k. \tag{33}$$

Provided that no state is absorbing, this model gives rise to recurring shifts in the parameters. Standard practice seems to be not to conduct much testing to identify the number of regimes, k, and many papers simply assume the presence of two states.

Again these models have been applied extensively in empirical analysis. Rene Garcia and Perron (1996) and Andrew Ang and Geert Bekaert (2002) use regime-switching models to capture the dynamics in U.S. interest rates, while Gabriel Perez-Quiros and Timmermann (2000) use these models to predict stock returns. Some papers find evidence that letting the state transition probabilities depend on forward-looking variables such as the leading indicator helps improve forecasting performance.

Perhaps surprisingly, relatively little work has been undertaking on merging the linear dynamic factor VARs with multivariate nonlinear specifications such as Markov switching models as proposed by Diebold and Glenn D. Rudebusch (1996). Some empirical findings have indicated the potential of this type of model (see, e.g., Marcelle Chauvet 1998 for an application to factor modeling and Massimo Guidolin and Timmermann (2006, forthcoming) in the context of multivariate regime switching models applied to forecast stock and bond returns and interest rates). Moreover, Markov Chain Monte Carlo methods which are useful for estimating these models are now widely available (Chang-Jin Kim and Charles R. Nelson 1999), so the application of these types of models to real time forecasting is less of a challenge than previously.

By accounting for model uncertainty, standard Bayesian methods are directly applicable for estimating the parameters of forecasting models even when these are time-varying. Furthermore, since the procedure is conditional on Z the fact that risk averages across sample information is incorporated through the prior, i.e., by the weighting of the relevant parameters.

As an example of a Bayesian analysis, Pesaran, Davide Pettenuzzo, and Timmermann (2006) propose a hidden Markov chain approach to forecast time series subject to multiple structural breaks. They assume a hierarchical prior setting in which breaks are viewed as shifts in an underlying Bernoulli process. The parameters of each regime are realizations of draws from a stationary meta distribution. Information about this distribution gets updated recursively through time as new breaks occur. This approach provides a way to forecast the frequency and size of future breaks. Their empirical findings for U.S. interest rates suggest that accounting for breaks in out-of-sample forecasts can be important, particularly at long forecast horizons.[9] Gary M. Koop and Simon M. Potter (forthcoming) also develop Bayesian methods for forecasting under breaks.

6.3 *Unit Roots*

Granger (1966) found that many macroeconomic data have a spectral peak near frequency zero and Nelson and Charles I. Plosser (1982) followed this up by showing that it was difficult to reject the presence of unit roots in many macroeconomic and financial data.

[9] Further complicating the picture is the fact that it may be difficult to separate the effects of seasonalities, nonlinearities and structural change. For a careful analysis of some of these issues, see Van Dijk, Birgit Strikholm, and Teräsvirta (2003), and Philip Hans Franses and Van Dijk (2005).

From the classical perspective, most of the literature has revolved around whether or not to impose unit roots. In the univariate model, this amounts to choosing between a model in levels or differences. In multivariate models, the problem of choosing levels or differences also raises the possibility that error correction terms can be included. When each individual series has a unit root but some of them are common, we know from the Granger representation theorem (Robert F. Engle and Granger 1987) that the full system can be represented as an error correction model.[10]

Imposing unit roots reduces risk when the true parameter is sufficiently close to this value but increases risk for more distant alternatives, with the effect dependent on the forecast horizon. Sample size is also important since estimates become more precise with more data swinging the balance in favor of less constrained estimation methods.

The most prominent alternative to OLS estimation is the pretest estimator which sets the estimate equal to one if the pretest fails to reject a unit root (since this is the null being tested) and otherwise selects the OLS estimate. Diebold and Kilian (2000) examine this method, which increases the gain from always imposing a unit root at the cost of doing worse on average than the OLS estimator when the coefficient is further away from a unit root.

Intuition is more complicated in multivariate models because of the higher dimensionality of the problem which results in a much broader set of trade-offs for the effect of parameter estimation on risk. Transitory dynamics can also affect both the magnitude and in some cases the sign of the results. Hence general results—whether analytical or through simulation—are difficult to arrive at.

The issue of unit roots or near nonstationarity also arises for the common linear forecasting model $\hat{y}_t = \hat{\beta}_0 + \hat{\beta}_1 z_{t-1}$ when the regressor, z_{t-1}, has a trend of unknown form. There are many examples of such models being applied. Forecasts of stock returns by highly persistent variables such as the dividend–price ratio or earnings–price ratio fit this situation, as does inflation forecasting using interest rate levels, or forecasting changes in exchange rates with the forward premium. While methods have been proposed for hypothesis testing in these models, there is not much evaluation of the effect on forecasting.

As in the unit root case, when the innovations to the regressor are correlated with the residuals of the forecasting equation, risk becomes a nonconstant function of the nuisance parameters describing the form of the persistence in the data such as the degree of persistence of z and the covariance between innovations to y and z. While most theoretical work has focused on testing $\beta_1 = 0$, little attention has been paid to designing good forecasting procedures or examining the trade-offs between possible forecasting methods.

7. *Forecast Evaluation*

As noted in the introduction, one of the major differences between standard econometric problems and the forecasting problem is that the researcher receives feedback on how well their forecast actually performed. Thus, when a central bank forecasts next-year output growth or inflation, the following year it is able to see how far off the forecasts were. Evaluating forecasting procedures in light of this new information generates a dynamic process through which a number of important issues arise.

Forecast evaluation usually comprises two separate, but related, tasks, namely (i) providing summary statistics for measuring the precision of past forecasts and (ii) testing optimality properties of the forecasts by means of a variety of diagnostics. The latter involves checking whether the conditions implied by an optimal forecast hold in a particular sample.

[10] See Boriss Siliverstovs, Tom Engsted, and Niels Haldrup (2004) for a related analysis.

If the loss function is known up to a finite set of unknown parameters, this is a straightforward process. From the forecaster's first order condition (3) the generalized forecast errors, $\mathcal{L}'(f,y,z)$ or $\mathcal{L}'$ for short, should themselves be unpredictable, i.e., follow a martingale difference sequence given all current information used to construct the forecasts.

The nature of this orthogonality condition will of course depend both on the shape of the forecaster's loss function and on the presumed data generating process underlying future values of Y used to calculate the conditional expectation $E_Y[\mathcal{L}'|Z]$. For example, under MSE loss the optimal forecast is, as we have seen, the conditional expectation of Y given all current information, Z, and the generalized forecast error is simply proportional to the forecast error. Forecast errors should therefore have zero mean, be serially uncorrelated and be unpredictable given all current information. These properties are particular to the MSE loss function and need not hold in general (Patton and Timmermann 2007a).

7.1 *Forecast Precision*

A variety of performance measures can be reported. It is common practice to use hold-out (out-of-sample) observations to obtain a measure of risk. The idea is to split the available sample into two pieces, a regression set of R observations and a subsequent prediction set of P observations. For each of the P observations in the hold-out set, we can employ the forecast procedure as if we were actually in the position of forecasting out of sample, constructing a sample of forecasts $f(z_t)$ for $t = R + 1,\ldots,R + P$. Since this is not actually done in real time, it is sometimes referred to as simulated or pseudo real time forecasting.

Three different updating schemes are commonly used to simulate real time forecasts: Recursive forecasts where all data up to time t are used in the construction of each forecast and the data expands as t increases; rolling forecasts which use only the most recent fixed interval of the past data so that the length of the data window remains the same as t increases; and fixed forecasts where only data up to R is used for the entire future.[11]

The sample analog of the risk for either procedure is simply the average loss, $P^{-1}\sum_{t=R+1}^{R+P}\mathcal{L}(f(z_t,\theta),y_{t+1},z_t)$. Such measures are routinely computed in Monte Carlo studies and in studies using real data. Real data sets the densities of both Y and Z to their empirically observed densities, and hence is generally viewed as more interesting.

Given the arbitrariness of the scale in most loss functions, the raw average loss number is difficult to interpret. However under MSE and MAE loss, the number can be directly interpreted. For MAE the loss function is in the scale of the units of the outcome variable, and hence a clear picture of the loss is immediate. For MSE, as with variances more generally, the square root of the outcome is reported so that it is in units of the outcome variable (root MSE or RMSE).

Forecast performance measures are estimates of the expected loss and hence are surrounded by sampling variability. West (1996) derives asymptotic representations for the sampling distribution of the average loss under quite general assumptions on the data, loss function, and forecasting method. He also provides asymptotic normal limiting results when forecasts are constructed recursively, using one of the aforementioned estimation windows. Under a number of technical conditions,[12] the average risk functions are consistent and asymptotically normally distributed with a covariance matrix that depends on the

[11] Fixed forecasts enable the theoretical simplification that the parameter estimates are based solely on data outside the period over which the data are averaged, though it would seem unlikely that this method is used in practice. The other two methods have the additional complication of the parameters being functions of the data in the forecasting period. Valentina Corradi and Swanson (2007) discuss the impact of estimation methods on the construction of critical values for predictive inference.

[12] The loss functions must be twice differentiable, estimators for $\hat{\theta}$ must be asymptotically linear, as well as mixing and moment assumptions on various functions of the data.

randomness of the out-of-sample observations and has additional terms reflecting the variation that arises through the forecasts' dependence on estimated parameters.

The results of West (1996) show that it is appropriate only in special cases to use the standard asymptotic variance covariance matrix that ignores randomness in the estimated parameters of the forecasting model. One situation is when the same loss function is employed for estimating the parameters θ and evaluating the forecast provided that the data is covariance stationary.[13] In this case, orthogonalities between the out of sample errors and the estimated model deliver the asymptotic equivalence. The most interesting case is the linear forecasting model used to minimize MSE loss for which standard errors can be computed as usual from the sequence of realized losses. Alternatively, if the estimation sample is large relative to the sample over which the forecasts are evaluated, then the additional variation due to estimating θ will be small since parameter estimates will be close to their true values and hence estimation is negligible asymptotically.

Some issues limit direct application of these results, however. When the hold-out sample either remains a fixed or a negligible proportion of the full sample, the coefficients of the forecasting models converge to their pseudo-true values. If two or more of the models are asymptotically equivalent (for example, if one nests another), then asymptotically the forecasts will be perfect correlated and the asymptotic approximation to the covariance matrix of the risks during the hold out sample is singular.

When choosing the measure in which to report forecasting results, it should be borne in mind that a forecast that may be good according to one measure (MSE), may not be good in terms of another measure, e.g., correctly predicted signs. To see this, consider the following simple example from Steve Satchell and Timmermann (1995):

$$y = f + \varepsilon,$$

where y is the outcome, f is the forecast, and ε is the forecast error which has standard deviation σ. If f and ε are independent, the probability of predicting the sign of y is a decreasing function of the mean squared prediction error, σ^2. Conversely, if f and ε are dependent, in general no such relationship between the mean squared prediction error and the probability of predicting the sign of y will hold. To see this, consider the 2×2 case

$$\begin{array}{ccc} \varepsilon \backslash f & -1 & 1 \\ -\sigma_1 & p_{11} & p_{12} \\ \sigma_2 & p_{21} & p_{22} \end{array}$$

where $\sigma_1 > \sigma_2 > 1$. Notice that

$$\Pr(sign(f) = sign(y)) = p_{11} + p_{22}$$

$$MSE(\varepsilon) = \sigma_1^2(p_{11} + p_{12}) + \sigma_2^2(p_{21} + p_{22}).$$

Now choose $0 < \delta < p_{ij}$ such that

$$\tilde{p}_{11} = p_{11} + 2\Delta$$

$$\tilde{p}_{12} = p_{12} - \Delta$$

$$\tilde{p}_{21} = p_{21} - \Delta$$

$$\tilde{p}_{22} = p_{22}$$

under these new probabilities we have

$$\Pr(sign(f) = sign(y)) = p_{11} + p_{22} + 2\Delta$$

$$MSE(\varepsilon) = \sigma_1^2(p_{11} + p_{12} + \Delta) + \sigma_2^2(p_{21} + p_{22} - \Delta).$$

We have thus increased the probability of correctly predicting the sign yet simultaneously increased the MSE. The general

[13] The latter condition may not hold when using real-time data, see Todd E. Clark and Michael W. McCracken (2007).

message from this simple example is that forecasting models with low MSE need not also be the ones with a high proportion of correctly predicted signs which is what may be most important in some applications.

7.2 *Efficiency Tests*

The hold-out sample can also be employed to approximate the first order condition through its sample equivalent

$$(34) \quad P^{-1} \sum_{t=R+1}^{R+P} \mathcal{L}'(f(z_t,\theta), y_{t+1}, z_t) = 0.$$

Moreover, realizations of $\mathcal{L}'(f(z_t,\theta), y_{t+1}, z_t)$ should also be uncorrelated with any information available at time t. Hence it is common to test the condition that

$$(35) \quad P^{-1} \sum_{t=R+1}^{R+P} \mathcal{L}'(f(z_t,\theta), y_{t+1}, z_t) v_t = 0,$$

where v_t is any function of $\{z_s\}_{s=1}^{t}$. Such a test can be conducted by regressing $\mathcal{L}'$ on v_t and testing that the OLS coefficients are zero. A particular function of v_t that is often employed is the forecast itself, which is a function of z_t and hence is a possible choice for v_t.

Under MSE loss, $\mathcal{L}'(f(z_t,\theta), y_{t+1}, z_t) \propto y_{t+1} - f(z_t,\theta) = e_{t+1}$ and thus is proportional to the forecast error. Hence (34) simply tests if the forecast errors have zero mean, and (35) tests that forecast errors are uncorrelated with any information available at the time that the forecast is made. For these reasons, (34) is known as an unbiasedness test and (35) is known as an orthogonality test. The most popular form of these tests is the Mincer–Zarnowitz (1969) regression

$$(36) \quad y_{t+1} = \beta_c + \beta f(z_t,\theta) + u_{t+1},$$

where u_{t+1} is an error satisfying $E[u_{t+1}|z_t] = 0$. Unbiasedness can now be tested through the joint constraint that $\beta_c = 0$ and $\beta = 1$.

Tests such as (34) and (35) examine whether or not the information in z_t has been used efficiently in the construction of the forecast. This is an important issue because a rejection of the test would suggest that improved forecasts are possible given the available data. It is also important from the perspective of testing rationality when the forecasts $f(z_t,\theta)$ are constructed by agents that are expected to be acting rationally and z_t is data that would have been available to those agents when they constructed their forecasts.

To examine these tests from an econometric perspective, recall that $f(z_t,\theta)$—and possibly also the instrument v_t—is constructed using parameter estimates based on data up to time t. When evaluating the sampling distribution for the regression estimates in the unbiasedness or orthogonality tests (34) and (35), we must therefore consider the sampling variability that arises through the fact that the variables in the regression are constructed. West and McCracken (1998) provide results for these regressions covering a number of methods for constructing the forecasts and v_t. Under assumptions similar to those in West (1996), they show that the coefficients in the regression tests are asymptotically normal, although the variance covariance matrix may need to be adjusted to allow for the additional variation arising from sampling variation in the underlying parameter estimates.

An additional practical concern involves the specification of v_t as a function of z_t. Often there are numerous candidate variables in z_t. This, combined with the possibility that we could use any functional form of z_t as an instrument, means that the list of candidates is practically unlimited. Any test of orthogonality has power only in the direction of the included instrument, v_t. For example, in forecasting inflation with v_t set to past interest rates, the test would be capable of picking up any additional explanatory power in interest rates but not for other variables. The same is true for getting the functional form correct. Avoiding the first problem—picking the wrong z_t to include—is difficult. For the second problem, Corradi and Swanson

(2002) suggest a nonparametric method for estimating a general function of the included elements of z_t.

For loss functions other than mean squared loss, $\mathcal{L}'(.)$ is no longer equivalent to the forecast errors. Hence it is possible that forecast errors are not mean zero and that past information may well be correlated with forecast errors even when the forecast is constructed optimally. Indeed, it is clear from (34) and (35) that the tests rely on the use of the correct loss function. Michael P. Keane and David E. Runkle (1990) write "If forecasters have differential costs of over- and underprediction, it could be rational for them to produce biased forecasts. If we were to find that forecasts are biased, it could still be claimed that forecasters were rational if it could be shown that they had such differential costs" (p. 719).

Rationality tests may thus reject, not because the forecaster is using information inefficiently but because the loss function has not been correctly specified. This is an important issue since the loss function is generally unknown even though it is invariably assumed to be of the MSE type. Elliott, Komunjer, and Timmermann (2005) examine a class of asymmetric quadratic loss functions

$$(37)\quad \mathcal{L}(e_{t+1};\alpha) \equiv [\alpha + (1 - 2\alpha)\cdot I(e_{t+1} < 0)]|e_{t+1}|^2,$$

where α $(0 < \alpha < 1)$ is the asymmetry parameter. This loss function reduces to MSE when $\alpha = 0.5$. Regressing forecast errors on v_t (as would be appropriate for MSE loss) results in coefficients on v_t that converge to the true coefficient plus an extra term $(1 - 2\alpha)E[v_t v_t']^{-1}E[v_t|e_{t+1}|]$. If v_t contains a constant term (which is usually the case), then $E[v_t|e_{t+1}|]$ is always nonzero and orthogonality tests based on MSE loss will reject asymptotically as a result of using a misspecified loss function.[14]

[14] See Bryan Campbell and Eric Ghysels (1995) for an interesting nonparametric evaluation of forecasts.

In general, future values of $\mathcal{L}'(.)$ should not themselves be predictable given any variables in the forecaster's current information set. A joint test of forecast efficiency (rationality) can thus readily be conducted within the context of a given family of loss functions which yields $\mathcal{L}'(.)$ as a function of a finite set of unknown parameters. If the test is rejected, either the forecaster did not use information efficiently or the family of loss functions was incorrectly specified.

In situations where the loss function is not known up to a small set of shape parameters, it is possible to use tests that trade off assumptions about the underlying data generating process against much weaker assumptions on the loss function (such as homogeneity properties). Patton and Timmermann (2007b) show that when loss is only required to be a homogenous function of the forecast error, while the data generating process can have dynamics in the first and second conditional moments (thus covering a large range of nonlinear-in-mean specifications, ARCH models, etc.), a simple quantile regression test can be used to test forecast optimality.

When several forecasts at multiple horizons are simultaneously available (as in the case with many survey forecasts), this offers significant advantages in terms of constructing tests for forecast efficiency that do not depend on knowing which information was available to the forecaster. Assuming that the forecaster makes efficient use of all historical information, under MSE loss and a stationary data generating process we have that $MSE_{h_L} > MSE_{h_S}$, where $h_L > h_S$ are long and short forecast horizons respectively.[15] To see this, suppose the bias is zero at all horizons and that the variance of the optimal two-period forecast is smaller than that of the optimal one-period forecast. In this case, the variance of last period's two-step-ahead forecast

[15] For other loss functions, Patton and Timmermann (2007a) prove that the expected loss at the longer horizons must be greater than or equal to the expected loss at short horizons.

must be smaller than the variance of the current one-period forecast, contradicting the assumption that the current one-period forecast was optimal in the first place. Hence, under appropriate stationarity assumptions, expected loss must be nondecreasing in the length of the forecast horizon.

7.3 *Survey and Real Time Forecasts*

Survey data provide an ideal way to test whether economic forecasters use information efficiently. In this regard, the empirical evidence has been mixed. Bryan W. Brown and Shlomo Maital (1981) studied average forecasts of U.S. GNP and rejected unbiasedness and efficiency in six-month growth predictions. Victor Zarnowitz (1985) found only weak evidence against efficiency for the average forecast of U.S. growth but stronger evidence against efficiency for individual forecasters. Batchelor and Pami Dua (1991) report little evidence that forecast errors were correlated with their own past values. In contrast, Anthony Davies and Kajal Lahiri (1995) found evidence that forecast efficiency was rejected for up to half of the survey participants in their panel analysis.

Survey data on inflation expectations is another area where efficiency tests have been conducted. Stephen Figlewski and Paul Wachtel (1981) analyze expectations of individual respondents and frequently reject forecast rationality under squared loss. Frederic S. Mishkin (1981) also rejects rationality of survey forecasts of inflation. Zarnowitz (1985) finds evidence of systematic forecast errors for U.S. inflation and rejects unbiasedness for more than half of the survey participants. Pesaran and Martin Weale (2006) provide an extensive review of the literature on survey data in forecasting.

The real-time nature of economic forecasting affects all stages of the forecasting process: models must be formulated, selected, and estimated in real time. Evidence of model break-down or misspecification must also be examined in real time. It is not clear, for example, what one can conclude from full-sample evidence of forecast inefficiency. Unless the inefficiency was detectable at an earlier stage of the sample using information that was available historically, it cannot be established that the forecaster acted irrationally. For example, under MSE loss, the forecast errors should be mean-zero conditional only on the available information (including the forecasting model) at the point the forecast was formulated and not conditional on full-sample information.[16]

Real-time considerations even pertain to the data vintage that was available at a given point in time and could have been used to formulate and evaluate a forecasting model. Dean Croushore (2006) and Croushore and Tom Stark (2003) make it clear that key macroeconomic data such as GDP growth are subject to important revisions, partly due to regular updates from preliminary to secondary and later data releases, partly due to changes in the methodology used to measure a particular variable. These revisions can lead not only to changes in the estimated parameters but can also affect the dynamic lag structure or functional form of the forecasting model and hence change conclusions regarding predictive relationships (Jeffery D. Amato and Swanson 2001). Data revisions are even more important for composite series such as the index of leading indicators whose composition may change due to past failures in forecasting (Diebold and Rudebusch 1991).

These points emphasize that it is important to use the original data vintages when simulating the real-time out-of-sample forecasting process and evaluating the precision of the resulting forecasts.

7.4 *Evaluation of Density Forecasts*

Forecasting is one case where "one size fits all" does not hold. Forecast users have different loss functions and therefore require different optimal point forecasts. It may

[16] See further discussion in Pesaran and Timmermann (2005a).

therefore be better to provide forecast densities instead of point forecasts. Many agencies now provide such information. For example, the Bank of England reports the "river of blood" forecast that shows their forecast of likely inflation outcomes by various shades of red. Similarly, the European Forecasting Network reports density forecasts for a range of macroeconomic variables. With such a density forecast in hand, decisionmakers with different loss functions will be able to separately solve for their optimal decision.

Two potential problems arise in comparing density forecasts to point forecasts. First, it is more difficult to estimate the whole density than to provide a single point forecast. Second, different estimation methods will be better for certain features of the density, and the loss function has information that is useful in suggesting which features of the density are important and which are not.

For the first of these, consider that for any lin–lin loss function the best estimator estimates the quantile of interest and not the entire density. Any density estimator (see, e.g., Anthony S. Tay and Kenneth F. Wallis 2000) may well trade off precision at the required quantile against precision over the entire density. Hence sample information in the density estimator may not be as good as that of the quantile estimator. Not all estimators are created equal.

For the second case, consider a binary outcome. The density for a binary outcome is equivalent to an estimate of a probability of the positive outcome, which in turn is simply a parameter estimate. Hence for this special case parameter estimation and density estimation are equivalent. In the case of a misspecified parametric density, Elliott and Lieli (2006) show that estimation that takes the loss function into account can provide a better estimate of the probability of a positive outcome for those loss functions. The estimator depends on the loss function. However each case yields a different estimate of the density. This is generally true for parametric estimation. Because parametric density estimation is the estimation of the parameters of the density, and different loss functions suggest different estimation techniques for the parameters, estimating the density without paying attention to the loss function and ultimate use of the forecast density involves estimation trade-offs (either implicit or explicit) that favor some users at the expense of others.

Although density forecasts are still not commonly reported, a literature has emerged on how such forecasts should be evaluated. A basic tool used to this end is the so-called probability integral transform. This is simply the inverse of the cumulative density function, F^{-1}, implied by a particular parametric forecasting model. When applied to the actual realization of the predicted variable (y), $F^{-1}(y)$ should be drawn from a uniform distribution and be independently and identically distributed over time. This argument ignores the effect of using estimated parameters of course, but this type of test has regained popularity following the study by Diebold, Todd A. Gunther, and Tay (1998). Corradi and Swanson (2006a, 2006b) cover these methods and provide a comprehensive summary of current tests in this area.

Bayesian methods provide the predictive density $f_Y(y|z) = \int f(y|z,\theta)\pi(\theta|z)\,d\theta$. When the outcome is realized, it can be compared to the density the model suggests it should be a draw from. A natural statistic to compute is the p-value of this outcome, y, i.e. $P(y < Y^p)$ where Y^p is the random variable with density $f_Y(y)$. If this p-value is extreme, it might bring the quality of the forecasting model into question. This evaluation method is used by, for example, Pesaran, Pettenuzzo, and Timmermann (2006) to assess the quality of forecasts of interest rates from various models.

Despite issues with estimation, there is one major advantage of the provision of a density forecast, especially when the decision maker and the forecaster are different. Density forecasts convey the uncertainty in the decision-making environment, in perhaps

a better way than expressions such as MSE do to decisionmakers. For an interesting example, see Charles H. Whiteman (1996) who recounts his experience with providing density forecasts to Iowa state officials.

8. *Comparing and Combining Forecasts*

Decisionmakers often have access to more than one forecast. When faced with multiple forecasts, two very different strategies are possible: to seek out the best single forecasting model or to attempt to combine forecasts generated across all or a subset of models. The first approach requires being able to formally compare the forecasting performance across several models, while the latter requires a method for estimating the weights on the models used in the combination. We cover both issues in this section.

One distinction that has not been important so far but becomes crucial in the context of forecast comparisons is the difference between forecasting models and forecasting methods. The former refer to a class of particular (parametric) specifications. Forecasting methods are a broader concept and comprise rules used to select a particular forecasting model at a given point in time as well as the approach used to estimate the forecasting model's parameters—e.g., rolling versus expanding windows.

8.1 *Forecast Comparisons*

Since there is typically considerable uncertainty over the forecasting model, often we observe a wide array of forecasts attempting to predict the same sequence of outcomes. This has led to a literature on comparing the performance of different forecasting approaches. The idea is to compare the risk of two or more forecasts in order to choose the one that is best. For forecasting procedures $f^i(z,\theta)$, $i = 1,\ldots,n$, this means trying to determine which of the risks $R(f^i(z,\theta),\theta)$ is smallest.

As with the evaluation of a single forecast, we can examine in-sample and out-of-sample performance. A hold-out sample can be used to construct an estimate of average risk, but with n different forecasting procedures this now becomes an $n \times 1$ vector of averages. These are then compared. Often the ordering of estimates of risk is examined, as well as the forecasting ability of general classes of models compared to some benchmark model.

A large number of papers aim to show that one particular forecasting model (e.g., a nonlinear specification) outperforms another benchmark model. However, it is difficult to extract any general rules from empirical studies in this literature since the best approach generally depends on the type of variable under consideration (i.e., nominal versus real data, data with a small or large persistent component), the data frequency and even the sample period.

More interesting from a general perspective are attempts to rank forecasting procedures over a wider range of data sets and see which ones perform well on average. Such an exercise is presented in Stock and Watson (1999a), who examine linear autoregressive models (with different subset selection methods such as AIC and BIC) along with commonly employed nonlinear models such as neural networks and LSTAR models across a large number of U.S. macroeconomic data series. Stock and Watson found empirically that LSTAR models on average were outperformed by neural network models, which in turn were outperformed (except at the one month horizon) by autoregressions.

Horse races between competing forecasting models abound in the empirical literature. Because of the presence of strong common components in many forecasts (often representing autoregressive dynamics in the predicted variable) and short, overlapping samples, forecasts produced by different models are often sufficiently close that it is not possible to distinguish between the models with much statistical precision (see, e.g., Timmermann 2007 for a comparison of the IMF's forecasts to private sector consensus forecasts).

From a practical point of view, three issues should determine which of the many available methods for forecast comparison to use, namely (i) the nature of the null hypothesis, i.e., a null that two (or more) models have identical risk when evaluated at their (pseudo-true) parameter values versus the null of equal predictive accuracy of the forecasting methods after averaging out random variations in parameter estimates; (ii) accounting for estimation uncertainty versus ignoring it; and (iii) whether the forecasting models are nested or not.

On the first point, most comparisons test the null hypothesis that all forecasting methods are equally precise in the sense that they have identical risk:

$$H_0 : R(f^1(z,\theta),\theta_1) = \cdots = R(f^n(z,\theta_n),\theta_n), \tag{38}$$

where $\theta_1,\ldots,\theta_n$ are the pseudo-true parameters under models $1,\ldots n$. This is the null hypothesis that West (1996) and many subsequent studies consider. This is akin to viewing forecasting performance as a specification test for the underlying models.

In contrast, Giacomini and White (2006) consider the comparison of forecasting methods that comprise not only the prediction model but also the estimation method and length of the estimation sample. The null they study is quite different from that in (38). Their null hypothesis is concerned with testing that the accuracy of all models is identical. This involves taking expectations over Y, Z, and the parameter estimates $\hat{\theta}_1,\ldots,\hat{\theta}_n$, which are random variables.

The Giacomini–White analysis shifts the focus away from comparisons based on average performance towards the conditional expectation of differences in performance across forecasting methods. One advantage of this approach is that it directly accounts for the effect of parameter uncertainty by expressing the null in terms of estimated parameters and estimation windows. Provided that estimation uncertainty does not vanish asymptotically, nested models can thus be compared under this approach. In practice, this means that the Giacomini–White approach is most relevant to the comparison of forecasting models estimated using rolling windows.

Turning to the second point, in an important paper that spurred many of the subsequent studies in the literature, Diebold and Roberto S. Mariano (1995) suggest using the standard t-statistic for testing equivalence in forecasting performance for pairwise model comparisons ($n = 2$) by taking the difference of the estimated losses and testing if the resulting time series has zero mean. For scaling, they suggest a robust estimator of the variance and suggest comparing this t-statistic to the standard normal distribution. Special cases of the West (1996) results are able to justify use of standard variance estimators for the Diebold and Mariano test.[17] Their approach does not, however, account for parameter estimation errors.

West (1996) is the first paper to account for the effect of parameter estimation error on forecast comparisons when forecasts are updated recursively through time. Clark and West (2004) provide an interesting illustration of how important parameter estimation can be in the comparison of a benchmark model with few or none parameters (e.g., the prevailing mean) versus a more heavily parameterized alternative model that may include time-varying predictor variables. Even when the larger model is true, because it involves estimation of more parameters and hence is more subject to parameter estimation error, we would expect this model to perform worse in finite samples than the simpler (biased) model, unless the predictive power of the extra regressor(s) is sufficiently large. Clark and West propose a test that

[17] The most prominent special case is when the expected value of the derivative of the loss function with respect to θ is zero evaluated at the true θ.

accounts for this problem by correcting for parameter estimation error.

Turning to the third and final point, a limitation of the methods developed in Diebold and Mariano (1995) and West (1996) is that they only apply when the forecasting models are non-nested. Clark and McCracken (2001) extend this work and develop tests for comparing nested models in the presence of parameter estimation uncertainty. This is the case most commonly faced by empirical researchers and thus is an important step forward in this area.

Whether the nested or non-nested case applies to a given forecast comparison can be surprisingly difficult to determine. In practice it is often forecasting methods as opposed to forecasting models that are being compared, whereas most theory is developed for comparing forecasting models. A given forecasting method, when applied recursively through time, may select different forecasting models at different points in time. This means that the models selected by two different forecasting methods sometimes could be nested while at other times could be non-nested. To our knowledge, no test exists at the present time that handles this complication, making forecast comparisons a tricky exercise in practice.

8.2 *Comparing Large Sets of Models*

If a model has good forecasting performance but is selected as the "best" model from a much larger set of candidate models, this could simply be due to random chance—or luck, rather than skill, as one might say if the "model" was a fund manager. After all, we would expect five out of one hundred randomly selected forecasting models to reject the null of no predictability if the size of our test is five percent. The practice of not choosing the forecasting model ex ante (based, say, on theoretical arguments), and only choosing the model after inspecting the test results for a large set of models is referred to as data snooping. It need not be a problem if the model's performance can be tested on a fresh sample as is true in many disciplines outside economics and finance. However, the model's forecasting performance on the original test sample will, by construction, be inflated and cannot be taken at face value.

When a large set of models needs to be compared ($n > 2$), the "data snooping" method of White (2000) can be employed. This method compares a set of risk estimates generated by a range of individual forecasts to the risk of a benchmark model. The null hypothesis is that the best of the forecasting methods is no better than the benchmark model

$$(39)\quad H_0 : \min_{i=1,\ldots,n} \left(R(f^i(z),\hat{\theta}_i) - R(f^b(z),\hat{\theta}_b)\right) = 0,$$

where $R(f^b(z),\hat{\theta}_b)$ is the benchmark performance. The alternative hypothesis is that the best of the forecast methods outperforms the benchmark, i.e. that it has lower risk:

$$(40)\quad H_1 : \min_{i=1,\ldots,n} \left(R(f^i(z),\hat{\theta}_i) - R(f^b(z),\hat{\theta}_b)\right) < 0.$$

Since the distribution of the risk differentials is asymptotically normal under the assumptions of the method (based on the results of West 1996), this amounts to constructing a test for the maximum of a set of joint normals with unknown covariance matrix. White solves this problem by employing a bootstrap procedure to the estimates of risk. This bypasses the need to compute the unknown variance covariance matrix and directly estimates the p-value for the test.

Peter Reinhard Hansen (2005) shows that when poor models are added to the set of candidate models, such that the benchmark is better than other models, the test proposed by White (2000) becomes increasingly conservative. He proposes a null distribution that accounts for the possibility that underperforming models may not be as good as the benchmark. Effectively this means that the null distribution becomes sample dependent and that the addition of a large number of underperforming models will not affect the asymptotic distribution of the test statistic. Finally, Hansen shows that basing the test on

normalized measures such as the student-t statistic can be very important in practice.

Controlling for data snooping can be important empirically. In the context of forecasting models for daily stock market returns based on technical trading rules, Ryan Sullivan, Timmermann, and White (1999) find that data snooping can account for what otherwise appears to be strong evidence of return predictability.

8.3 *Forecast Encompassing*

Yock Y. Chong and Hendry (1986) introduced the idea of forecast encompassing, which can be applied when choosing between forecasting models. Under MSE loss, the idea is similar to the orthogonality regressions (35) although the additional information v_t is no longer a subset of z_t but instead consists of forecasts or forecast errors from other forecasting methods. The idea is simple: If other forecasts have information relevant for the predicted variable that is not contained in the original forecast, then such forecasts will enter the orthogonality regression with a nonzero weight. This would mean that the original forecast did not include all relevant information. Conversely, if orthogonality holds, then the first forecast is said to encompass the other forecasts because it incorporates all the relevant information that the other forecasts have.

Christina D. Romer and David H. Romer (2000) provide an interesting comparison of the Federal Reserve Green Book inflation forecasts with private sector forecasts using encompassing regressions. Assuming MSE loss, they find evidence that the Fed inflation forecasts encompass the private sector forecasts. This conclusion is questioned by Capistran (2006) who finds evidence of significant biases of opposite sign in the Fed's forecasts during the pre- and post-Volker periods. Averaged over the full sample, the bias is small, but this conceals evidence of a tendency to underpredict inflation in the pre-Volker sample followed by subsequent overpredictions.

To test if a particular forecast (null model) encompasses a set of alternative forecasts, a regression of the forecast error from the null model on the difference between the other forecast errors and that of the null model (e_t^*) can be undertaken using a t- or an F- test (see, e.g., Clements and Hendry 1998, p. 265)

$$(41) \qquad e_t^* = \beta_0 + \beta_1(e_t^1 - e_t^*) + \cdots + \beta_n(e_t^n - e_t^*) + \varepsilon_t.$$

Alternative forms of this test have been suggested by David I. Harvey, Stephen J. Leybourne, and Newbold (1998) and Clark and McCracken (2001) when two forecasts are being compared. To handle the problem that forecast errors depend on the estimated parameter, θ, the results of West and McCracken (1998) can be used. When the models are nested, the singularity of the joint distribution of the forecast errors is again a problem. Clark and McCracken (2001) show that in these cases the asymptotic distribution of a rescaled statistic can be approximated with a function of Brownian motions and hence the distribution is nonstandard in this case.

8.4 *In-Sample versus Out-of-Sample Forecast Comparison*

The emphasis in the forecast comparison literature has been on out-of-sample comparisons and the goal has been to select the best forecasting model for a given loss function. However this problem can be recast to ask "which model is better?" This is a problem that has been closely examined in econometrics. For at least some model comparisons, tests conducted on the entire sample rather than on an artificially extracted hold-out sample might therefore be more appropriate and powerful (Inoue and Kilian 2004).

Which strategy to adopt for sample selection very much depends on the purpose of the analysis. If interest lies in testing implications of economic theories related to the presence of predictability in population, it is

best and most powerful to use the full sample. In contrast, if interest lies in testing for the presence of real time predictability under the conditions facing actual forecasters in finite samples, then the use of a hold-out sample may make sense. In the latter case, the bias-variance trade-off may benefit small, misspecified models even though these models do not have good population properties.

Under MSE loss, the problem of comparing forecast models reduces to the question of which forecast procedure is closest to the conditional expectation. This can be tested in the full sample without problems of nesting of the models so long as the data are sufficiently stationary and not too dependent, although such tests are of course subject to the earlier mentioned caveats.[18]

The desire to test out-of-sample forecasting performance is closely related to the uncertainty about the underlying data generating process and also a concern for the effect of any pretesting that might have occurred in constructing the forecasting models. Inoue and Kilian (2004) have questioned the practice of using a hold-out sample altogether arguing that it does not protect against data mining because the information available in "pseudo real time" experiments is the same as that available to someone with access to the full sample.

8.5 *Forecast Combinations*

Forecast comparisons are not meant as tools for selecting forecasting models, unless of course a decision rule based on the outcome of such tests is specified in advance. For example, if the null that two models have equal predictive accuracy is not rejected, a natural follow-up question is whether to use the forecasts from the first model, the second model or perhaps both. Even if one model seems to be better than another, it is not clear that it is optimal to ignore the forecasts from the weaker model altogether. The literature on forecast combinations attempts to address such issues.

Despite the many attempts to choose a single forecasting model, empirically it seems that combining forecasts from multiple models often outperforms forecasts from a single model. Clemen (1989) reviews the literature and finds that combinations outperform individual models in a wide range of forecasting problems. Spyros Makridakis and Michele Hibon (2000) find similar results involving the forecasting of 3003 data series. For U.S. macroeconomic series, Stock and Watson (1999a) find that combining the forecasts from several methods on average performed better than simply relying on forecasts from individual models such as neural networks, LSTAR or autoregressions. Marcellino (2004) reports similar results for European data.

An argument often used to justify forecast combinations is that they diversify against model uncertainty. Forecasting models are best viewed as simple approximations to a more complicated, and constantly evolving, reality. We would therefore expect them to be misspecified in many regards—for example, they may exclude important information that is not easily modeled or they may not adjust sufficiently fast to evidence of model breakdown. Some models may adapt very quickly to a change in the behavior of the predicted variable, while others adapt more slowly. To some extent forecast combination therefore provides insurance against "breaks" or other nonstationarities that may occur in the future.

Because it is not known a priori how and whether the world will change in the future, a sensible strategy is to combine the forecasts from two or more approaches. A key issue is to what extent different forecasts diversify against modeling risk—which depends on the correlation in forecast errors across models—and how much weight to assign to the various forecasts.

A direct answer to the question of how to obtain a set of combination weights is provided by J. M. Bates and Granger (1969) who suggest simply regressing the predicted

[18] Such assumptions are of course also required in the tests proposed by West (1996) and for results built on this paper.

variable y on the individual forecasts $f_i(z,\theta)$ along with a constant

$$(42) \qquad y = \beta_0 + \sum_{i=1}^{n} \beta_i f_i(z,\theta) + \varepsilon.$$

When the individual forecasts are believed to be unbiased, it is common to omit the intercept term and restrict the slope coefficients to sum to one in which case they can be interpreted as forecast combination weights. This approach assumes MSE loss but has been generalized to other loss functions and method of moment type estimators (Elliott and Timmermann 2004).

A comparison of the combination approach to encompassing explains why combining may be expected to be a more reasonable approach than selecting a single forecast, unless of course the true model is known to be included in the set of models under consideration and can be identified in practice. A single forecast only gets selected when the combination puts full weight on one of the forecasting methods while the rest are given zero weights. This is precisely the case where the forecast with a weight of one encompasses the other forecasts. However, this is a special case of the general concept of forecast combination, and so might be expected to be less commonly supported empirically than more evenly distributed weights.

In practice, although empirical evidence suggests that forecast combinations tend to outperform forecasts from a single model, strategies designed to obtain optimal combination weights are often outperformed by simple measures such as averaging the raw forecasts (i.e., giving all forecasts equal weights) or a trimmed set of these. If the models use roughly the same data sources and empirical techniques so differences in the performance across forecasting models are too small to be easily rejected by the data, they will tend to have similar error variances and covariances. In this situation, giving each forecast identical weights can be relatively efficient. Franz C. Palm and Zellner (1992) suggest other reasons—e.g., instability of the covariance between forecast errors or estimation error in the combination weights.

Which combination methodology is best may well depend on the state of the economy because the speed with which different forecasts incorporate shifts in the economy could vary. For example, when the economy is running at a normal pace, time-series models may provide the most accurate forecast because they make efficient use of historical information. However, these models may be slower at capturing or predicting turning points—such as the emergence of a recession—compared with seasoned professional forecasters with access to a much larger information set. This idea is consistent with findings reported in Elliott and Timmermann (2005) who use a regime switching approach to track variations in the forecasting performance of time-series and survey forecasts of six key macroeconomic variables and form a combined forecast.

Bayesian approaches to forecast combination are becoming increasingly popular in empirical studies. Bayesian Model Averaging has been proposed by, inter alia, Edward E. Leamer (1978) and Adrian E. Raftery, David Madigan, and Jennifer A. Hoeting (1997). Under this approach, the predictive density can be computed by averaging over a set of models, $M_i, i = 1,\ldots,n$:

$$(43) \qquad p_Y(y|z) = \sum_{i=1}^{n} p(M_i|z) p_Y(y|M_i,z).$$

Here $p(M_i|z)$ is the posterior probability of model M_i obtained from the model priors $\pi(M_i)$, the priors for the unknown parameters of each model, $\pi(\theta_i|M_i)$, and the likelihood of the models under consideration. $p_Y(y|M_i,z)$ is the predictive density of y under the ith model, M_i, given z obtained after uncertainty about the parameters θ_i has been integrated out using the posterior. Unlike the weights used in the classical least-squares combination literature, these weights do not account for correlations between

forecasts and the weights are always confined to the zero-unity interval. Palm and Zellner (1992) develop a general Bayesian framework for combinations where individual forecasts can be biased and the covariance matrix of the forecast errors may be unknown. More details are provided in Timmermann (2006) and Geweke and Whiteman (2006).

9. *Empirical Application*

To illustrate many of the issues discussed above, we consider the predictability of U.S. inflation and stock returns. For inflation, we use log first differences of the CPI while stock returns are captured by the value-weighted portfolio of U.S. stocks tracked by the Center for Research in Security Prices. Both series are measured at the monthly frequency and the sample period is 1959:1–2003:12. To initialize our parameter estimates, we use data from 1959:1 to 1969:12. We then generate out-of-sample forecasts from 1970:01 to 2003:12. Parameter estimates are either updated recursively, expanding the estimation window by one observation each month, or by means of a ten-year rolling window. Only data up to the previous month is therefore used to estimate the model parameters and generate forecasts for the current month. This is commonly referred to as a pseudo out-of-sample forecasting exercise.

We consider twelve forecasting approaches. The first is an autoregressive (AR) model

$$y_{t+1} = \beta_0 + \sum_{j=1}^{k} \beta_j y_{t+1-j} + \varepsilon_{t+1}, \tag{44}$$

where k is selected to minimize the BIC with a maximum of 18 lags and ε_{t+1} here and in subsequent models is regarded as white noise. The second model is a factor augmented AR model, using up to five common factors:

$$y_{t+1} = \beta_0 + \sum_{j=1}^{k} \beta_j y_{t+1-j} + \sum_{j=1}^{q} \gamma_j \psi_{j,t} + \varepsilon_{t+1}, \tag{45}$$

where $\psi_{j,t}$ is the jth factor and k and q are again selected to minimize the BIC (with $k \leq 18$ and $q \leq 5$). Factors are obtained using the principal components approach of Stock and Watson (2002) to a cross-section of 131 macroeconomic time series which begin in 1960. The factors are extracted in (simulated) real time using either a recursive or a rolling ten-year estimation window.

The third and fourth models are Bayesian VARs (BVARs) fitted to the variable of interest (inflation or stock returns) and the five factors:

$$z_{t+1} = \beta_0 + \sum_{j=1}^{k} \beta_j z_{t+1-j} + \varepsilon_{t+1}. \tag{46}$$

Here $z_t = (y_t, \psi_{1t}, \ldots, \psi_{5t})'$ and we include the most recent six months lags, i.e. $k = 6$. Following Litterman, own-lag terms at lag j have a prior variance of $0.04/j^2$, while off-diagonal lags have a prior variance of $0.0004/j^2$. Both a random walk prior and a white noise prior are considered. Under the random walk prior, the autoregressive parameters are shrunk toward unity, while under the white noise prior they are shrunk toward zero. Clearly the random walk prior is reasonable for the inflation example while the white noise prior is more reasonable for stock returns. We report both for each example to show the effect of the differences in prior choice.

Turning to the nonlinear specifications, we consider two logistic STAR models of the form

$$y_{t+1} = \theta_1' \eta_t + d_t \theta_{21}' \eta_t + \varepsilon_{t+1}, \tag{47}$$

where

$$\eta_t = (1, y_t)'$$

$$d_t = \begin{cases} 1/(1 + exp(\gamma_0 + \gamma_1 y_{t-3})) \\ 1/(1 + exp(\gamma_0 + \gamma_1 (y_t - y_{t-6}))) \end{cases}.$$

We refer to these as STAR1 and STAR2, respectively.

As more flexible nonlinear alternatives, we consider a single layer neural net model

$$(48) \quad y_{t+1} = \theta_0'\eta_t + \sum_{i=1}^{n}\theta_i g(\beta_i'\eta_t) + \varepsilon_{t+1}$$

with two hidden units ($n = 2$) as well as a two-layer neural net model

$$(49) \quad y_{t+1} = \theta_1'\eta_t + \sum_{i=1}^{n_2}\theta_{ig}\Big(\sum_{j=1}^{n_1}\beta_j g(\alpha_j'\eta_t)\Big) + \varepsilon_{t+1}$$

with two hidden units in the first layer ($n_1 = 2$) and one hidden layer in the second layer ($n_2 = 1$). For both neural net models, g is the logistic function and $\eta_t = (1, y_t, y_{t-1}, y_{t-2})$. Estimation uses search methods since α_j enters nonlinearly.

We also consider more traditional time-series forecasting methods such as exponential smoothing where the forecast f_t is generated by the recursion

$$(50) \quad f_{t+1} = \alpha f_t + (1 - \alpha)y_t,$$

subject to the initial condition that $f_1 = y_1$, and double exponential smoothing

$$(51) \quad f_{t+1} = \alpha(f_t + \lambda_{t-1}) + (1 - \alpha)y_t$$
$$\lambda_t = \beta(f_{t+1} - f_t) + (1 - \beta)\lambda_{t-1},$$

where $f_1 = 0, f_2 = y_2$ and $\lambda_2 = (y_2 - y_1)$. Here α and α and β, respectively, are determined so as to minimize the sum of squared forecast errors in real time.[19]

We finally consider a forecast combination approach that simply uses the equal-weighted average in addition to a very different approach that, at each point in time, selects the forecasting model with the best track record up to the present time and then uses this to generate a forecast for the following period.

[19] For a comprehensive treatment of exponential smoothing methods, see Hyndman et al. (2008).

In all cases, we apply the following "insanity filter" that constrains outlier forecasts: If the predicted change in the underlying variable is greater than any of the historical changes up to a given point in time, the forecast is replaced with a "no change" forecast.

Results in the form of out-of-sample, annualized root mean squared forecast errors (computed by multiplying the monthly RMSE values by the square root of 12) are presented in table 1. First consider the results under recursive parameter estimation. For inflation, the best model is the average forecast followed by exponential smoothing, the previous best model, the simple and factor-augmented AR models and the two-layer neural net model. Slightly worse forecasts are generated by the one-layer neural net and the BVARs, while the STAR models generate somewhat worse performances.

Overall, these results indicate that there is not much to differentiate between a cluster of the best forecasting models. This point is reinforced by the plots of predicted values from three of the models shown in figure 1. Inflation forecasts from seemingly very different approaches are quite similar and dominated by a persistent common component.

Turning to the stock returns and focusing again on the results under recursive estimation, table 1 shows that the best overall performance is delivered by the combined forecast and the simple and factor-augmented AR models. Once again the BVAR models perform rather poorly as do the STAR models and single layer neural nets. For stock returns that are not dominated by a strongly persistent component, there is more to differentiate between the time-series of forecasts as shown in figure 2. Overall, however, while a few approaches perform quite poorly it is difficult to distinguish with statistical precision between the forecasting performance among a cluster of reasonable forecasting models.

Forecast precision tends to deteriorate significantly for the BVAR and double exponential smoothing forecasts under the ten-year rolling estimation window. This happens both

TABLE 1
OUT-OF-SAMPLE FORECASTING PERFORMANCE (ANNUALIZED ROOT MEAN SQUARED ERROR) FOR VARIOUS FORECASTING MODELS, 1970–2003

	Inflation		S & P 500 Returns	
Model Name	Expanding window	10-year rolling window	Expanding window	10-year rolling window
Autoregressive (AR)	0.78	0.77	15.9	15.9
Factor-augmented AR	0.78	0.80	15.9	15.9
BVAR-random walk prior	0.81	0.92	17.7	20.0
BVAR-white noise prior	0.81	0.94	17.4	19.1
Exponential smoothing	0.76	0.76	16.0	16.3
Double exp. smoothing	0.78	0.83	16.2	18.6
STAR 1	0.88	0.80	16.8	17.5
STAR 2	0.83	0.81	17.0	17.4
One layer neural net	0.80	0.82	17.1	17.4
Two layer neural net	0.78	0.77	16.0	17.5
Combined forecast (average)	0.75	0.75	15.9	16.3
Previous best	0.77	0.78	16.1	16.5

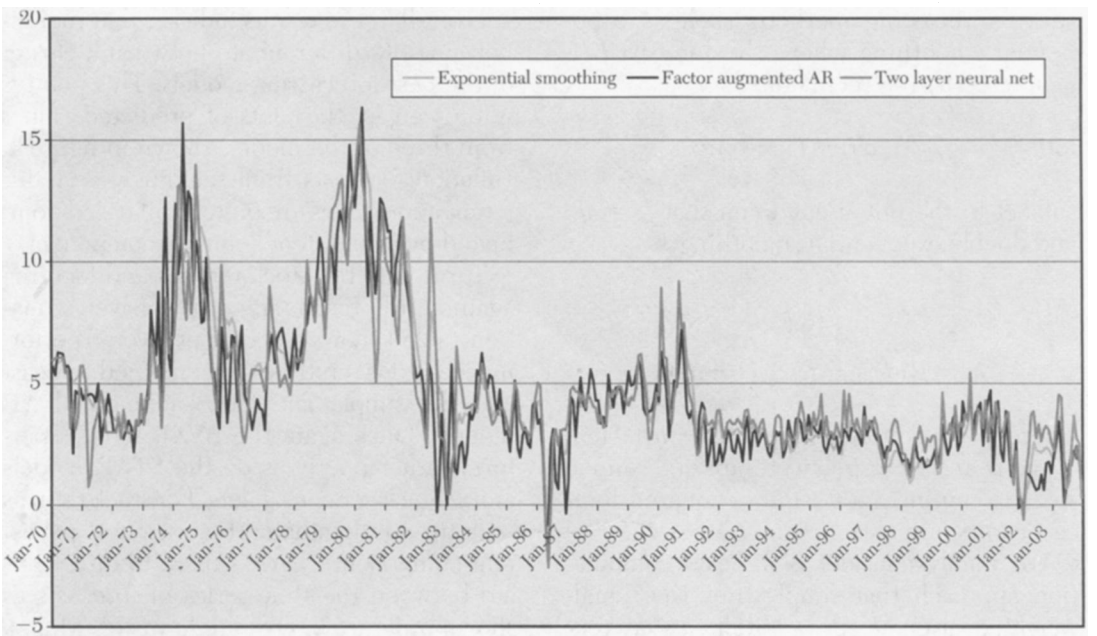

Figure 1. Inflation Forecasts

for inflation and stock returns. This deterioration is likely due to the larger estimation error associated with using a shorter estimation window, although one should not forget that there is a trade-off in the form of faster adaptability as witnessed by the improved forecasting performance observed for the first STAR model's inflation forecasts.

Evaluation of the forecasts from these models is complicated because some of them

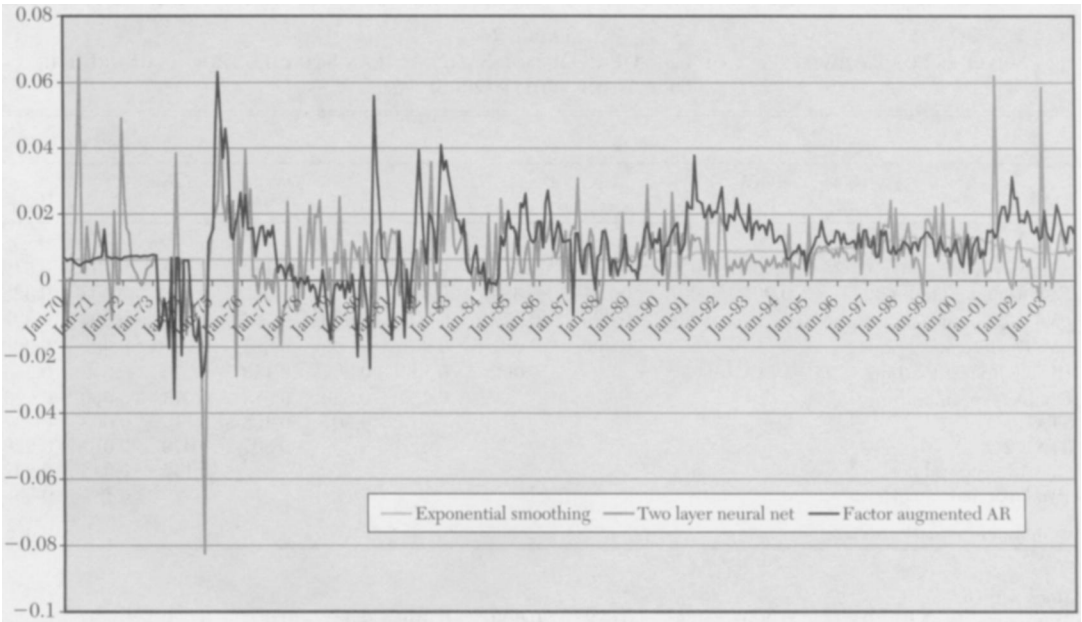

Figure 2. Forecasts of Stock Returns

are pair-wise nested (for example, the STAR and neural net models nest the AR models), while others are not (e.g., the factor-augmented AR models are not nested by the neural net models). As a diagnostic test, we simply compare the MSE performance of pairs of models using the Giacomini–White (2006) approach. Our results assume a rolling estimation window corresponding to ten years of monthly observations, i.e., 120 data points and therefore reflects the RMSE values reported in columns two and four in table 1.

Results from these pairwise comparisons are reported in table 2. While the BVAR and double exponential smoothing inflation forecasts are soundly rejected against those produced by the better models, for most of the other comparisons these tests do not have sufficient power to choose one model over another. The results are somewhat different in the case of the stock returns which, unlike the inflation series, do not contain a large persistent component and hence are more difficult to predict. There is little evidence to distinguish between the simple and factor-augmented AR models, the exponential smoothing, average and previous best forecasts of stock returns. Conversely, the BVARs, double exponential smoothing, STAR and neural net forecasts are generally rejected against the first group of forecasts. Parsimony seems to be key to successfully predict stock returns, particularly when a relatively short rolling estimation window of 120 observations is used.

Once again it is clear that although a few approaches perform very poorly and can be rejected out of hand, it is difficult to systematically differentiate between many of the other approaches.

The discussion has so far assumed MSE loss. Theory suggests that the form of the loss function alters the optimal functional form of the forecasting model. To illustrate this, we next generated forecasts under lin–lin loss, setting $p = 1$ and α equal to 0.35, 0.50 or 0.65 in equation (12), and considering either

TABLE 2
P-VALUES FOR PAIRWISE TEST OF IDENTICAL OUT-OF-SAMPLE MEAN SQUARED ERRORS UNDER THE GIACOMINI–WHITE (2006) TEST

	Model two										
Model one	Factor AR	BVAR -RW	BVAR -WN	Exp. Smoothing	Double exp. Smoothing	STAR_1	STAR_2	One layer NN	Two layer NN	Average	Best
Inflation											
Autoregressive (AR)	0.028	0.000	0.000	0.813	0.012	0.155	0.075	0.142	0.943	0.162	0.542
Factor-augmented AR		0.005	0.001	0.124	0.324	0.806	0.674	0.569	0.133	0.002	0.328
BVAR-random walk prior			0.226	0.000	0.027	0.003	0.008	0.047	0.001	0.000	0.001
BVAR-white noise prior				0.000	0.013	0.001	0.003	0.021	0.000	0.000	0.000
Exponential smoothing					0.000	0.125	0.084	0.128	0.766	0.347	0.364
Double exp. smoothing						0.224	0.659	0.913	0.015	0.000	0.020
STAR 1							0.507	0.496	0.115	0.003	0.427
STAR 2								0.807	0.079	0.010	0.220
One layer neural net									0.132	0.049	0.242
Two layer neural net										0.167	0.588
Combined forecast (average)											0.076
Stock returns											
Autoregressive (AR)	0.937	0.000	0.000	0.073	0.000	0.010	0.001	0.009	0.060	0.150	0.177
Factor-augmented AR		0.000	0.000	0.050	0.000	0.004	0.000	0.004	0.047	0.054	0.155
BVAR-random walk prior			0.009	0.000	0.065	0.001	0.000	0.000	0.007	0.000	0.000
BVAR-white noise prior				0.000	0.528	0.048	0.022	0.034	0.100	0.000	0.002
Exponential smoothing					0.000	0.015	0.005	0.024	0.120	0.063	0.509
Double exp. smoothing						0.059	0.072	0.038	0.185	0.000	0.000
STAR 1							0.928	0.901	0.968	0.009	0.008
STAR 2								0.982	0.922	0.005	0.038
One layer neural net									0.904	0.007	0.017
Two layer neural net										0.098	0.176
Combined forecast (average)											0.528

the AR model or the factor-augmented AR model.[20] All forecasts were generated using a recursive estimation window. Results from this analysis are presented in table 3. For inflation the average value of the lin–lin loss function under the simple AR model is generally significantly below the values produced under the factor-augmented AR specification. This holds irrespective of which quantile is being considered. In contrast, for stock returns the two models produce almost identical out-of-sample forecasting performance.

These empirical results support many of the themes of our theoretical analysis. First, forecasts from seemingly very different approaches (e.g., linear versus nonlinear models) often produce very similar results—witness the similar RMSE performance of the neural nets and the autoregressive models. In part, these similarities arise because we truncate the forecasts from the nonlinear models when these are too far away from the historical sample data.[21] In other cases,

[20] These forecasts were generated using quantile regression, see Koenker and Bassett (1978). This already presents a nonlinear optimization problem so we only consider linear quantile specifications in our analysis.

[21] When extreme forecasts are not truncated, the RMSE for the stock return forecasts rises to 50 and 32 under the one- and two-layer neural net models, respectively. These values are three times and twice as large as the values reported in table 1.

TABLE 3
FORECASTING PERFORMANCE UNDER LIN-LIN LOSS,
FORECAST MODELS ESTIMATED BY QUANTILE REGRESSION

Model	Quantile	Inflation	Stock return
Autoregressive	0.35	0.27	5.72
Factor-augmented AR	0.35	0.34	5.71
Autoregressive	0.5	0.30	6.09
Factor-augmented AR	0.5	0.40	6.07
Autoregressive	0.65	0.29	5.64
Factor-augmented AR	0.65	0.42	5.68

nonlinear models can generate poor forecasts due to their sensitivity to outliers and their imprecisely estimated parameters. This last point is illustrated through the performance of the STAR models which generally was quite poor.

Secondly, it is difficult to outperform simple approaches such as a parsimonious autoregressive model. Simple forecasting approaches tend to generate relatively smooth and stable forecasts without being subject to too much parameter estimation error.

Third, and as an extension of the previous point, it appears that in many cases there are only marginal gains (in terms of out-of-sample RMSE performance) over and above projection on past values of the series themselves from considering the additional information that can be extracted from large data sets. For persistent variables such as inflation, a linear autoregressive component is clearly the single most important predictive component, while for stock returns it is difficult to come up with predictor variables with significant predictive value.

Fourth, our results support the finding that forecast combination offers an attractive approach for many economic and financial variables. The average forecast produced the best or second-best performance among all approaches for both inflation and stock returns. Thus, while forecast combination does not always generate the single best performance, it usually beats most alternatives unless some extremely poor models have been left in the mix of models that get combined.

Fifth, the loss function clearly matters in practice. We saw that, under MSE loss, the purely autoregressive and factor-augmented autoregressive models produced essentially indistinguishable forecasting performance. In contrast, under lin–lin loss, the simple autoregressive forecasts were better for the inflation series, although they were nearly identical in the case of the stock returns.

Finally, model instability and/or sensitivity of forecasting performance to the sample period is clearly an issue. Table 1 compares the MSE performance under a recursive estimation approach which uses an expanding estimation window against that of a rolling ten-year window that can better accommodate shifts in the underlying data generating process. In many cases, the choice of estimation window makes a sizeable difference. If estimation error was the predominant effect, we would expect the ten-year rolling forecasts uniformly to be worse than the forecasts based on the expanding estimation window. This is exactly what we find for stock returns where there is no evidence that shortening the estimation window leads to improvements in any of the models. For inflation, however, we see that for half of the models the out-of-sample forecasting performance

either improves or stays the same as a result of going from the expanding to the rolling estimation window. Indeed, in the case of the first STAR model, the latter approach produced substantially better forecasts.

10. *Conclusion*

The menu of forecasting methodologies available to the applied economist has expanded vastly over the last few decades. No single approach is currently dominant and choice of forecasting method is often dictated by the situation at hand such as the forecast user's particular needs, data availability, and expertise in experimenting with different classes of models and estimation methods. Economic forecasts are often only one piece of information used in conjunction with a decisionmaker's prior beliefs and other sources of information. Moreover, such forecasts are often used as a way to assign different weights on various possible scenarios. Purely statistical approaches based on complicated "black box" approaches have with few exceptions so far failed to generate much attention among economists and are not used to the extent one might otherwise have expected.

Although the situation is still evolving, recent research in the forecasting literature has supported some broad conclusions:

- Careful attention to the forecaster's objectives is important not only in the forecast evaluation stage but also in the estimation and model selection stages. For example, if the forecaster's loss function suggests that overpredictions are more costly than underpredictions (or vice versa) and a particular quantile of the forecast distribution best summarizes the economic objectives of the forecasting exercise, then quantile rather than least squares estimation should be used;
- Models of economic and financial time series are often found to be unstable through time and so forecasting models are best viewed as approximations or tracking devices. As a consequence, one should not expect that the same forecasting model will continue to dominate in different historical samples;
- Choice of the sample period used to estimate the parameters of the forecasting model is therefore important. Using the longest possible data sample or a simple rolling window is not necessarily the best approach if more precise information about the cause of model instability is available (e.g., institutional shifts, changes in tax policy or legislation, large technology or supply shocks). Since the nature and form of model instability may often not be very clear, more research is required to design robust forecasting approaches that detect and incorporate model instability in a variety of situations;
- Forecast combination has often been found to offer an attractive alternative to the approach of seeking to identify a single best forecasting model. In part this stems from the fact that combination allows forecasters to hedge against model uncertainty and shifts in models' (relative) forecasting performance;
- Overfitting is an overriding concern in forecasting because of the short time series often encountered and the difficulty in getting independent data samples that can be used to cross-validate the forecasting models. This problem is exacerbated for financial time series where the signal-to-noise ratio tends to be very low. Parameter estimation error also is the likely reason why including additional economic variables in a forecasting model, which may seem justified ex ante, often fails to lead to the expected improvement in terms of out-of-sample forecasting performance;
- It is often difficult to distinguish with much statistical precision between the forecasts generated by seemingly very different forecasting methods. When large differences in forecasting performance occur, this often has to do with

the tendency of nonlinear forecasting models to generate outliers in the forecast error distribution due to their sensitivity to the particular sample used for parameter estimation. How such outliers are dealt with then becomes important in practice;

- Guidance from economic theory is important at several stages of the forecasting process. Besides assisting in the choice of the forecaster's objective function, economic theory can be helpful in selecting categories of variables to be considered as potential predictors and in imposing long-run restrictions that may reduce parameter estimation error. Econometric methods can then be used for variable selection among the predictors deemed potentially relevant from a theoretical perspective (often a large set), for specification of the short-run dynamics, and for determination of the functional form of the forecasting model.

References

Amato, Jeffery D., and Norman R. Swanson. 2001. "The Real-Time Predictive Content of Money for Output." *Journal of Monetary Economics*, 48(1): 3–24.

Andersen, Torben G., Tim Bollerslev, Peter F. Christoffersen, and Francis X. Diebold. 2006. "Volatility and Correlation Forecasting." In *Handbook of Economic Forecasting*, ed. G. Elliott, C. W. J. Granger, and A. Timmermann. Amsterdam: Elsevier, North-Holland, 777–878.

Ang, Andrew, and Geert Bekaert. 2002. "Regime Switches in Interest Rates." *Journal of Business and Economic Statistics*, 20(2): 163–82.

Artis, Michael, and Massimiliano Marcellino. 2001. "Fiscal Forecasting: The Track Record of the IMF, OECD and EC." *Econometrics Journal*, 4(1): S20–36.

Bai, Jushan. 1997. "Estimation of a Change Point in Multiple Regression Models." *Review of Economics and Statistics*, 79(4): 551–63.

Bai, Jushan, and Pierre Perron. 1998. "Estimating and Testing Linear Models with Multiple Structural Changes." *Econometrica*, 66(1): 47–78.

Batchelor, Roy, and Pami Dua. 1991. "Blue Chip Rationality Tests." *Journal of Money, Credit, and Banking*, 23(4): 692–705.

Batchelor, Roy, and David A. Peel. 1998. "Rationality Testing under Asymmetric Loss." *Economics Letters*, 61(1): 49–54.

Bates, J. M., and Clive W. J. Granger. 1969. "The Combination of Forecasts." *Operations Research Quarterly*, 20(4): 451–68.

Box, George E. P., and Gwilym M. Jenkins. 1970. *Time Series Analysis: Forecasting and Control*. San Francisco: Holden-Day.

Brayton, F., and P. Tinsley. 1996. "A Guide to FRB/US: A Macroeconomic Model of the United States." Board of Governors of the Federal Reserve System, Finance and Economics Discussion Series, no. 96-42.

Breiman, Leo. 1995. "Better Subset Regression Using the Nonnegative Garrote." *Technometrics*, 37(4): 373–84.

Breiman, Leo. 1996. "Bagging Predictors." *Machine Learning*, 24(2): 123–40.

Brown, Bryan W., and Shlomo Maital. 1981. "What Do Economists Know? An Empirical Study of Experts' Expectations." *Econometrica*, 49(2): 491–504.

Campbell, Bryan, and Eric Ghysels. 1995. "Federal Budget Projections: A Nonparametric Assessment of Bias and Efficiency." *Review of Economics and Statistics*, 77(1): 17–31.

Capistran, Carlos Carmona. 2006. "Bias in Federal Reserve Inflation Forecasts: Is the Federal Reserve Irrational or Just Cautious?" Banco de Mexico Working Paper, no.2006-14.

Chauvet, Marcelle. 1998. "An Econometric Characterization of Business Cycle Dynamics with Factor Structure and Regime Switching." *International Economic Review*, 39(4): 969–96.

Chong, Yock Y., and David F. Hendry. 1986. "Econometric Evaluation of Linear Macro-economic Models." *Review of Economic Studies*, 53(4): 671–90.

Christoffersen, Peter F., and Francis X. Diebold. 1997. "Optimal Prediction under Asymmetric Loss." *Econometric Theory*, 13(6): 808–17.

Christoffersen, Peter F., and Francis X. Diebold. 2006. "Financial Asset Returns, Direction-of-Change Forecasting, and Volatility Dynamics." *Management Science*, 52(8):1273–87.

Christoffersen, Peter F., and Kris Jacobs. 2004. "The Importance of the Loss Function in Option Valuation." *Journal of Financial Economics*, 72(2): 291–318.

Clark, Todd E., and Michael W. McCracken. 2001. "Tests of Equal Forecast Accuracy and Encompassing for Nested Models." *Journal of Econometrics*, 105(1): 85–110.

Clark, Todd E., and Michael W. McCracken. 2007. "Tests of Equal Predictive Ability with Real-Time Data." Federal Reserve Bank of Kansas City, Research Working Paper, no. RWP07-06.

Clark, Todd E., and Kenneth D. West. 2004. "Using Out-of-Sample Mean Squared Prediction Errors to Test the Martingale Difference Hypothesis." Federal Reserve Bank of Kansas City, Research Working Paper, no. RWP 04-03.

Clemen, Robert T. 1989. "Combining Forecasts: A Review and Annotated Bibliography." *International Journal of Forecasting*, 5(4): 559–83.

Clements, Michael P., and David F. Hendry. 1998. *Forecasting Economic Time Series*. Cambridge; New York and Melbourne: Cambridge University Press.

Clements, Michael P., and David F. Hendry. 2002.

"Modelling Methodology and Forecast Failure." *Econometrics Journal*, 5(2): 319–44.
Clements, Michael P., and David F. Hendry. 2006. "Forecasting with Breaks." In *Handbook of Economic Forecasting*, ed. G. Elliott, C. W. J. Granger, and A. Timmermann. Amsterdam: Elsevier, North-Holland, 605–57.
Corradi, Valentina, and Norman R. Swanson. 2002. "A Consistent Test for Nonlinear Out of Sample Predictive Accuracy." *Journal of Econometrics*, 110(2): 353–81.
Corradi, Valentina, and Norman R. Swanson. 2006a. "Predictive Density and Conditional Confidence Interval Accuracy Tests." *Journal of Econometrics*, 135(1–2): 187–228.
Corradi, Valentina, and Norman R. Swanson. 2006b. "Predictive Density Evaluation." In *Handbook of Economic Forecasting*, ed. G. Elliott, C. W. J. Granger, and A. Timmermann. Amsterdam: Elsevier, North-Holland, 197–284.
Corradi, Valentina, and Norman R. Swanson. 2007. "Nonparametric Bootstrap Procedures for Predictive Inference Based on Recursive Estimation Schemes." *International Economic Review*, 48(1): 67–109.
Croushore, Dean. 2006. "Forecasting with Real-Time Macroeconomic Data." In *Handbook of Economic Forecasting*, ed. G. Elliott, C. W. J. Granger, and A. Timmermann. Amsterdam: Elsevier, North-Holland, 961–82.
Croushore, Dean, and Tom Stark. 2003. "A Real-Time Data Set for Macroeconomists: Does the Data Vintage Matter?" *Review of Economics and Statistics*, 85(3): 605–17.
Davies, Anthony, and Kajal Lahiri. 1995. "A New Framework for Analyzing Survey Forecasts Using Three-Dimensional Panel Data." *Journal of Econometrics*, 68(1): 205–27.
Del Negro, Marco, Frank Schorfheide, Frank Smets, and Rafael Wouters. 2007. "On the Fit of New Keynesian Models." *Journal of Business and Economic Statistics*, 25(2):123–62.
Diebold, Francis X., Todd A. Gunther, and Anthony S. Tay. 1998. "Evaluating Density Forecasts with Applications to Financial Risk Management." *International Economic Review*, 39(4): 863–83.
Diebold, Francis X., and Lutz Kilian. 2000. "Unit-Root Tests Are Useful for Selecting Forecasting Models." *Journal of Business and Economic Statistics*, 18(3): 265–73.
Diebold, Francis X., and Roberto S. Mariano. 1995. "Comparing Predictive Accuracy." *Journal of Business and Economic Statistics*, 13(3): 253–63.
Diebold, Francis X., and Glenn D. Rudebusch. 1991. "Forecasting Output with the Composite Leading Index: A Real-Time Analysis." *Journal of the American Statistical Association*, 86(415): 603–10.
Diebold, Francis X., and Glenn D. Rudebusch. 1996. "Measuring Business Cycles: A Modern Perspective." *Review of Economics and Statistics*, 78(1): 67–77.
Doan, Thomas, Robert B. Litterman, and Christopher A. Sims. 1984. "Forecasting and Conditional Projection Using Realistic Prior Distributions." *Econometric Reviews*, 3(1): 1–100.
Ehrbeck, Tilman, and Robert Waldmann. 1996. "Why Are Professional Forecasters Biased? Agency versus Behavioral Explanations." *Quarterly Journal of Economics*, 111(1): 21–40.
Elliott, Graham. 2005. "Forecasting in the Presence of a Break." Unpublished.
Elliott, Graham, Ivana Komunjer, and Allan Timmermann. 2005. "Estimation and Testing of Forecast Rationality under Flexible Loss." *Review of Economic Studies*, 72(4): 1107–25.
Elliott, Graham, Ivana Komunjer, and Allan Timmermann. 2006. "Biases in Macroeconomic Forecasts: Irrationality or Asymmetric Loss?" Unpublished.
Elliott, Graham, Robert P. Lieli. 2006. "Predicting Binary Outcomes." Unpublished.
Elliott, Graham, and Ulrich K. Muller. 2006. "Efficient Tests for General Persistent Time Variation in Regression Coefficients." *Review of Economic Studies*, 73(4): 907–40.
Elliott, Graham, and Allan Timmermann. 2004. "Optimal Forecast Combinations under General Loss Functions and Forecast Error Distributions." *Journal of Econometrics*, 122(1): 47–79.
Elliott, Graham, and Allan Timmermann. 2005. "Optimal Forecast Combination under Regime Switching." *International Economic Review*, 46(4): 1081–1102.
Engle, Robert F., and Clive W. J. Granger. 1987. "Cointegration and Error Correction: Representation, Estimation, and Testing." *Econometrica*, 55(2): 251–76.
Figlewski, Stephen, and Paul Wachtel. 1981. "The Formation of Inflationary Expectations." *Review of Economics and Statistics*, 63(1): 1–10.
Forni, Mario, Marc Hallin, Marco Lippi, and Lucrezia Reichlin. 2000. "The Generalized Dynamic-Factor Model: Identification and Estimation." *Review of Economics and Statistics*, 82(4): 540–54.
Forni, Mario, Marc Hallin, Marco Lippi, and Lucrezia Reichlin. 2002. *The Generalized Dynamic Factor Model: One-Sided Estimation and Forecasting.* CEPR Discussion Paper, no. 3432.
Franses, Philip Hans, and Dick Van Dijk. 2005. "The Forecasting Performance of Various Models for Seasonality and Nonlinearity for Quarterly Industrial Production." *International Journal of Forecasting*, 21(1): 87–102.
Gallant, A. Ronald. 1981. "On the Bias in Flexible Functional Forms and an Essentially Unbiased Form: The Fourier Flexible Form." *Journal of Econometrics*, 15(2): 211–45.
Garcia, Rene, and Pierre Perron. 1996. "An Analysis of the Real Interest Rate under Regime Shifts." *Review of Economics and Statistics*, 78(1): 111–25.
Geweke, John. 2005. *Contemporary Bayesian Econometrics and Statistics.* Hoboken, N.J.: Wiley.
Geweke, John, and Charles H. Whiteman. 2006. "Bayesian Forecasting." In *Handbook of Economic Forecasting*, ed. G. Elliott, C. W. J. Granger, and A. Timmermann. Amsterdam: Elsevier, North-Holland, 3–80.
Giacomini, Raffaella, and Halbert White. 2006. "Tests of Conditional Predictive Ability." *Econometrica*, 74(6): 1545–78.

Granger, Clive W. J. 1966. "The Typical Spectral Shape of an Economic Variable." *Econometrica*, 34(1): 150–61.

Granger, Clive W. J. 1969. "Prediction with a Generalized Cost of Error Function." *Operations Research*, 20(2): 199–207.

Granger, Clive W. J. 1999. "Outline of Forecast Theory Using Generalized Cost Functions." *Spanish Economic Review*, 1(2): 161–73.

Granger, Clive W. J., and Mark J. Machina. 2006. "Forecasting and Decision Theory." In *Handbook of Economic Forecasting*, ed. G. Elliott, C. W. J. Granger, and A. Timmermann. Amsterdam: Elsevier, North-Holland, 81–98.

Granger, Clive W. J., and Paul Newbold. 1986. *Forecasting Economic Time Series*. Second edition. Orlando, Fla.; London; Sydney and Toronto: Harcourt, Brace, Jovanovich; Academic Press.

Granger, Clive W. J., and M. Hashem Pesaran. 2000. "Economic and Statistical Measures of Forecast Accuracy." *Journal of Forecasting*, 19(7): 537–60.

Guidolin, Massimo, and Allan Timmermann. 2006. "An Econometric Model of Nonlinear Dynamics in the Joint Distribution of Stock and Bond Returns." *Journal of Applied Econometrics*, 21(1): 1–22.

Guidolin, Massimo, and Allan Timmermann. Forthcoming. "Forecasts of US Short-Term Interest Rates: A Flexible Forecast Combination Approach." *Journal of Econometrics*.

Hamilton, James D. 1989. "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle." *Econometrica*, 57(2): 357–84.

Hansen, Peter Reinhard. 2005. "A Test for Superior Predictive Ability." *Journal of Business and Economic Statistics*, 23(4): 365–80.

Harvey, Andrew. 2006. "Forecasting with Unobserved Components Time Series Models." In *Handbook of Economic Forecasting*, ed. G. Elliott, C. W. J. Granger, and A. Timmermann. Amsterdam: Elsevier, North-Holland, 327–412.

Harvey, Andrew, and Siem Jan Koopman. 1993. "Forecasting Hourly Electricity Demand Using Time-Varying Splines." *Journal of the American Statistical Association*, 88(424):1228–36.

Harvey, David I., Stephen J. Leybourne, and Paul Newbold. 1998. "Tests for Forecast Encompassing." *Journal of Business and Economic Statistics*, 16(2): 254–59.

Hendry, David F., and Hans-Martin Krolzig. 2004. "Automatic Model Selection: A New Instrument for Social Science." *Electoral Studies*, 23(3): 525–44.

Hong, Harrison, and Jeffrey D. Kubik. 2003. "Analyzing the Analysts: Career Concerns and Biased Earnings Forecasts." *Journal of Finance*, 58(1): 313–51.

Hyndman, Rob J., Anne B. Koehler, J. Keith Ord, and Ralph D. Snyder. 2008. *Forecasting with Exponential Smoothing: The State Space Approach*. New York: Springer.

Inoue, Atsushi, and Lutz Kilian. 2004. "In-Sample or Out-of-Sample Tests of Predictability: Which One Should We Use?" *Econometric Reviews*, 23(4): 371–402.

Inoue, Atsushi, and Lutz Kilian. 2006. "On the Selection of Forecasting Models." *Journal of Econometrics*, 130(2): 273–306.

Inoue, Atsushi, and Lutz Kilian. Forthcoming. "How Useful is Bagging in Forecasting Economic Time Series? A Case Study of U.S. CPI Inflation." *Journal of the American Statistical Association*.

Ito, Takatoshi. 1990. "Foreign Exchange Rate Expectations: Micro Survey Data." *American Economic Review*, 80(3): 434–49.

Jagannathan, Ravi, and Tongshu Ma. 2003. "Risk Reduction in Large Portfolios: Why Imposing the Wrong Constraints Helps." *Journal of Finance*, 58(4): 1651–83.

James, W., and Charles Stein. 1961. "Estimation with Quadratic Loss." In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1*, ed. J. Neyman. Berkeley and Los Angeles: University of California Press, 361–79.

Kadiyala, K. Rao, and Sune Karlsson. 1993. "Forecasting with Generalized Bayesian Vector Autoregressions." *Journal of Forecasting*, 12(3–4): 365–78.

Kadiyala, K. Rao, and Sune Karlsson. 1997. "Numerical Methods for Estimation and Inference in Bayesian VAR-Models." *Journal of Applied Econometrics*, 12(2): 99–132.

Keane, Michael P., and David E. Runkle. 1990. "Testing the Rationality of Price Forecasts: New Evidence from Panel Data." *American Economic Review*, 80(4): 714–35.

Kilian, Lutz. 1999. "Exchange Rates and Monetary Fundamentals: What Do We Learn from Long-Horizon Regressions?" *Journal of Applied Econometrics*, 14(5): 491–510.

Kilian, Lutz and Simone Manganelli. 2006. "The Central Banker as a Risk Manager: Estimating the Federal Reserve's Preferences under Greenspan." Unpublished.

Kim, Chang-Jin, and Charles R. Nelson. 1999. *State-Space Models with Regime Switching: Classical and Gibbs-Sampling Approaches with Applications*. Cambridge and London: MIT Press.

Koenker, Roger W., and Gilbert Bassett. 1978. "Regression Quantiles." *Econometrica*, 46(1): 33–50.

Koop, Gary M., and Simon M. Potter. Forthcoming. "Forecasting and Estimating Multiple Change-Point Models with an Unknown Number of Change-Points." *Review of Economic Studies*.

Leamer, Edward E. 1978. *Specification Searches: Ad Hoc Inference with Nonexperimental Data*. New York: Wiley.

Ledoit, Olivier, and Michael Wolf. 2003. "Improved Estimation of the Covariance Matrix of Stock Returns with an Application to Portfolio Selection." *Journal of Empirical Finance*, 10(5): 603–21.

Leitch, Gordon, and J. Ernest Tanner. 1991. "Economic Forecast Evaluation: Profits versus the Conventional Error Measures." *American Economic Review*, 81(3): 580–90.

Lim, Terence. 2001. "Rationality and Analysts' Forecast Bias." *Journal of Finance*, 56(1): 369–85.

Litterman, Robert B. 1980. "A Bayesian Procedure for Forecasting with Vector Autoregressions."

Massachusetts Institute of Technology Working Paper.
Litterman, Robert B. 1986. "Forecasting with Bayesian Vector Autoregressions—Five Years of Experience." *Journal of Business and Economic Statistics*, 4(1): 25–38.
Lopez, Jose A., and Christian A. Walter. 2001. "Evaluating Covariance Matrix Forecasts in a Value-at-Risk Framework." *Journal of Risk*, 3(3): 69–97.
Ludvigson, Sydney C., and Serena Ng. 2005. "Macro Factors in Bond Risk Premia." Unpublished.
Ludvigson, Sydney C., and Serena Ng. 2007. "The Empirical Risk–Return Relation: A Factor Analysis Approach." *Journal of Financial Economics*, 83(1): 171–222.
Makridakis, Spyros, and Michele Hibon. 2000. "The M3-Competition: Results, Conclusions and Implications." *International Journal of Forecasting*, 16(4): 451–76.
Mamaysky, Harry, Matthew Spiegel, and Hong Zhang. 2007. "Improved Forecasting of Mutual Fund Alphas and Betas." *Review of Finance*, 11(3): 359–400.
Marcellino, Massimiliano. 2004. "Forecast Pooling for European Macroeconomic Variables." *Oxford Bulletin of Economics and Statistics*, 66(1): 91–112.
Marcellino, Massimiliano, James H. Stock, and Mark W. Watson. 2006. "A Comparison of Direct and Iterated Multistep AR Methods for Forecasting Macroeconomic Time Series." *Journal of Econometrics*, 135(1–2): 499–526.
Meese, Richard A., and Kenneth Rogoff. 1983. "Empirical Exchange Rate Models of the Seventies: Do They Fit Out of Sample?" *Journal of International Economics*, 14(1–2): 3–24.
Miller, Alan. 2002. *Subset Selection in Regression*. Second edition. Boca Raton: Chapman and Hall.
Mincer, Jacob, and Victor Zarnowitz. 1969. "The Evaluation of Economic Forecasts." In *Economic Forecasts and Expectations: Analysis of Forecasting Behavior and Performance*, ed. J. Mincer. New York: National Bureau of Economic Research, 14–20.
Mishkin, Frederic S. 1981. "Are Market Forecasts Rational?" *American Economic Review*, 71(3): 295–306.
Nelson, Charles R., and Charles I. Plosser. 1982. "Trends and Random Walks in Macroeconomic Time Series: Some Evidence and Implications." *Journal of Monetary Economics*, 10(2): 139–62.
Newey, Whitney K., and James L. Powell. 1987. "Asymmetric Least Squares Estimation and Testing." *Econometrica*, 55(4): 819–47.
Ottaviani, Marco, and Peter Norman Sorensen. 2006. "The Strategy of Professional Forecasting." *Journal of Financial Economics*, 81(2): 441–66.
Pagan, Adrian. 2003. "Report on Modelling and Forecasting at the Bank of England." *Bank of England Quarterly Bulletin*, 43(1): 60–88.
Palm, Franz C., and Arnold Zellner. 1992. "To Combine or Not to Combine? Issues of Combining Forecasts." *Journal of Forecasting*, 11(8): 687–701.
Patton, Andrew J., and Allan Timmermann. 2007a. "Properties of Optimal Forecasts under Asymmetric Loss and Nonlinearity." *Journal of Econometrics*, 140(2): 884–918.
Patton, Andrew J., and Allan Timmermann. 2007b. "Testing Forecast Optimality under Unknown Loss." *Journal of the American Statistical Association*, 102: 1172–84.
Paye, Bradley S., and Allan Timmermann. 2006. "Instability of Return Prediction Models." *Journal of Empirical Finance*, 13(3): 274–315.
Perez-Quiros, Gabriel, and Allan Timmermann. 2000. "Firm Size and Cyclical Variations in Stock Returns." *Journal of Finance*, 55(3): 1229–62.
Pesaran, M. Hashem, Davide Pettenuzzo, and Allan Timmermann. 2006. "Forecasting Time Series Subject to Multiple Structural Breaks." *Review of Economic Studies*, 73(4):1057–84.
Pesaran, M. Hashem, and Spyros Skouras. 2002. "Decision-Based Methods for Forecast Evaluation." In *A Companion to Economic Forecasting*, ed. M. Clements and D. Hendry. Malden, Mass. and Oxford: Blackwell, 241–67
Pesaran, M. Hashem, and Allan Timmermann. 2005a. "Real-Time Econometrics." *Econometric Theory*, 21(1): 212–31.
Pesaran, M. Hashem, and Allan Timmermann. 2005b. "Small Sample Properties of Forecasts from Autoregressive Models under Structural Breaks." *Journal of Econometrics*, 129(1–2): 183–217.
Pesaran, M. Hashem, and Allan Timmermann. 2007. "Selection of Estimation Window in the Presence of Breaks." *Journal of Econometrics*, 137(1): 134–61.
Pesaran, M. Hashem, and Martin Weale. 2006. "Survey Expectations." In *Handbook of Economic Forecasting*, ed. G. Elliott, C. W. J. Granger, and A. Timmermann. Amsterdam: Elsevier, North-Holland, 715–76.
Psaradakis, Zacharias, and Fabio Spagnolo. 2005. "Forecast Performance of Nonlinear Error-Correction Models with Multiple Regimes." *Journal of Forecasting*, 24(2): 119–38.
Racine, Jeffrey. 2001. "On the Nonlinear Predictability of Stock Returns Using Financial and Economic Variables." *Journal of Business and Economic Statistics*, 19(3): 380–82.
Raftery, Adrian E., David Madigan, and Jennifer A. Hoeting. 1997. "Bayesian Model Averaging for Linear Regression Models." *Journal of the American Statistical Association*, 92(437): 179–91.
Rapach, David E., and Mark E. Wohar. 2006. "Structural Breaks and Predictive Regression Models of Aggregate U.S. Stock Returns." *Journal of Financial Econometrics*, 4(2): 238–74.
Robertson, John C., and Ellis W. Tallman. 1999. "Vector Autoregressions: Forecasting and Reality." *Federal Reserve Bank of Atlanta Economic Review*, 84(1): 4–18.
Romer, Christina D., and David H. Romer. 2000. "Federal Reserve Information and the Behavior of Interest Rates." *American Economic Review*, 90(3): 429–57.
Rossi, Barbara. 2006. "Are Exchange Rates Really Random Walks? Some Evidence Robust to Parameter Instability." *Macroeconomic Dynamics*, 10(1): 20–38.
Rossi, Barbara, and Raffaella Giacomini. 2006. "Detecting and Predicting Forecast Breakdowns."

Duke University, Department of Economics Working Paper, no. 06-01.
Sarno, Lucio, and Mark P. Taylor. 2002. "Purchasing Power Parity and the Real Exchange Rate." *IMF Staff Papers*, 49(1): 65–105.
Satchell, Steve, and Allan Timmermann. 1995. "An Assessment of the Economic Value of Non-linear Foreign Exchange Rate Forecasts." *Journal of Forecasting*, 14(6): 477–97.
Scharfstein, David S., and Jeremy C. Stein. 1990. "Herd Behavior and Investment." *American Economic Review*, 80(3): 465–79.
Schorfheide, Frank. 2005. "VAR Forecasting under Misspecification." *Journal of Econometrics*, 128(1): 99–136.
Siliverstovs, Boriss, Tom Engsted, and Niels Haldrup. 2004. "Long-Run Forecasting in Multicointegrated Systems." *Journal of Forecasting*, 23(5): 315–35.
Sims, Christopher A. 1980. "Macroeconomics and Reality." *Econometrica*, 48(1): 1–48.
Sims, Christopher A. 2002. "The Role of Models and Probabilities in the Monetary Policy Process." *Brookings Papers on Economic Activity*, 2: 1–40.
Skouras, Spyros. 2001. "Decisionmetrics: A Decision-Based Approach to Econometric Modeling." Santa Fe Institute Working Paper, no. 01-11-064.
Stock, James H., and Mark W. Watson. 1996. "Evidence on Structural Instability in Macroeconomic Time Series Relations." *Journal of Business and Economic Statistics*, 14(1):11–30.
Stock, James H., and Mark W. Watson. 1998. "Median Unbiased Estimation of Coefficient Variance in a Time-Varying Parameter Model." *Journal of the American Statistical Association*, 93(441): 349–58.
Stock, James H., and Mark W. Watson. 1999a. "A Comparison of Linear and Nonlinear Univariate Models for Forecasting Macroeconomic Time Series." In *Cointegration, Causality, and Forecasting: A Festschrift in Honour of Clive W. J. Granger*, ed. R. F. Engle and H. White. Oxford and New York: Oxford University Press, 1–44.
Stock, James H., and Mark W. Watson. 1999b. "Forecasting Inflation." *Journal of Monetary Economics*, 44(2): 293–335.
Stock, James H., and Mark W. Watson. 2002. "Macroeconomic Forecasting Using Diffusion Indexes." *Journal of Business and Economic Statistics*, 20(2): 147–62.
Stock, James H., and Mark W. Watson. 2005. "An Empirical Comparison of Methods for Forecasting Using Many Predictors." Unpublished.
Sullivan, Ryan, Allan Timmermann, and Halbert White. 1999. "Data-Snooping, Technical Trading Rule Performance, and the Bootstrap." *Journal of Finance*, 54(5): 1647–91.
Svensson, Lars E. O. 1997. "Inflation Forecast Targeting: Implementing and Monitoring Inflation Targets." *European Economic Review*, 41(6): 1111–46.
Swanson, Norman R., and Halbert 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(3): 265–75.
Tay, Anthony S., and Kenneth F. Wallis. 2000. "Density Forecasting: A Survey." *Journal of Forecasting*, 19(4): 235–54.
Teräsvirta, Timo. 2006. "Forecasting Economic Variables with Nonlinear Models." In *Handbook of Economic Forecasting*, ed. G. Elliott, C. W. J. Granger, and A. Timmermann. Amsterdam: Elsevier, North-Holland, 413–57.
Teräsvirta, Timo, Dick Van Dijk, and Marcelo C. Medeiros. 2005. "Linear Models, Smooth Transition Autoregressions, and Neural Networks for Forecasting Macroeconomic Time Series: A Re-examination." *International Journal of Forecasting*, 21(4): 755–74.
Tibshirani, Robert. 1996. "Regression Shrinkage and Selection via the Lasso." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 58(1): 267–88.
Timmermann, Allan. 2006. "Forecast Combinations." In *Handbook of Economic Forecasting*, ed. G. Elliott, C. W. J. Granger, and A. Timmermann. Amsterdam: Elsevier, North-Holland, 135–96.
Timmermann, Allan. 2007. "An Evaluation of the World Economic Outlook Forecasts." *IMF Staff Papers*, 54(1): 1–33.
Trueman, B. 1994. "Analyst Forecasts and Herding Behavior." *Review of Financial Studies*, 7(1): 97–124.
Van Dijk, Dick, Birgit Strikholm, and Timo Teräsvirta. 2003. "The Effects of Institutional and Technological Change and Business Cycle Fluctuations on Seasonal Patterns in Quarterly Industrial Production Series." *Econometrics Journal*, 6(1): 79–98.
Varian, Hal R. 1974. "A Bayesian Approach to Real Estate Assessment." In *Studies in Bayesian Econometrics and Statistics in Honor of Leonard J. Savage*, ed. S. E. Fienberg and A. Zellner. Amsterdam: North-Holland, 195–208.
Waggoner, Daniel F., and Tao Zha. 1999. "Conditional Forecasts in Dynamic Multivariate Models." *Review of Economics and Statistics*, 81(4): 639–51.
Weiss, Andrew A. 1996. "Estimating Time Series Models Using the Relevant Cost Function." *Journal of Applied Econometrics*, 11(5): 539–60.
West, Kenneth D. 1996. "Asymptotic Inference about Predictive Ability." *Econometrica*, 64(5): 1067–84.
West, Kenneth D., Hali J. Edison, and Dongchul Cho. 1993. "A Utility-Based Comparison of Some Models of Exchange Rate Volatility." *Journal of International Economics*, 35(1–2):23–45.
West, Kenneth D., and Michael W. McCracken. 1998. "Regression-Based Tests of Predictive Ability." *International Economic Review*, 39(4): 817–40.
West, Mike, and Jeff Harrison. 1997. *Bayesian Forecasting and Dynamic Models*. Second edition. New York and Heidelberg: Springer.
White, Halbert. 2000. "A Reality Check for Data Snooping." *Econometrica*, 68(5):1097–1126.
White, Halbert. 2001. *Asymptotic Theory for Econometricians*. Second edition. San Diego; London and Tokyo: Academic Press.
Whiteman, Charles H. 1996. "Bayesian Prediction under Asymmetric Linear Loss: Forecasting State Tax Revenues in Iowa." In *Forecasting, Prediction and Modeling in Statistics and Econometrics:*

Bayesian and Non-Bayesian Approaches, ed. W. O. Johnson, J. C. Lee, and A. Zellner. New York: Springer, 149–68.

Zarnowitz, Victor. 1985. "Rational Expectations and Macroeconomic Forecasts." *Journal of Business and Economic Statistics*, 3(4): 293–311.

Zellner, Arnold. 1986. "Bayesian Estimation and Prediction Using Asymmetric Loss Functions." *Journal of the American Statistical Association*, 81(394): 446–51.

Zellner, Arnold, and Chansik Hong. 1989. "Forecasting International Growth Rates Using Bayesian Shrinkage and Other Procedures." *Journal of Econometrics*, 40(1): 183–202.