

# Testing Forecast Optimality Under Unknown Loss

Andrew J. PATTON and Allan TIMMERMANN

Empirical tests of forecast optimality have traditionally been conducted under the assumption of mean squared error loss or some other known loss function. In this article we establish new testable properties that hold when the forecaster's loss function is unknown but testable restrictions can be imposed on the data-generating process, trading off conditions on the data-generating process against conditions on the loss function. We propose flexible estimation of the forecaster's loss function in situations where the loss depends not only on the forecast error, but also on other state variables, such as the level of the target variable. We apply our results to the problem of evaluating the Federal Reserve's forecasts of output growth. Forecast optimality is rejected if the Fed's loss depends only on the forecast error. However, the empirical findings are consistent with forecast optimality provided that overpredictions of output growth are costlier to the Fed than underpredictions, particularly during periods of low economic growth.

KEY WORDS: Federal Reserve; Forecast efficiency; Forecast evaluation; Loss function; Output growth.

## 1. INTRODUCTION

Forecasts are extensively used to guide the decisions of individuals, businesses, and macroeconomic policy-makers; thus what constitutes a "good" or optimal forecast is of great practical relevance. Indeed, knowledge of the properties of an optimal forecast has been used, *inter alia*, in tests of the efficient market hypothesis in financial markets and in tests of the rationality of decision makers in various economic applications. Almost without exception, empirical work has relied on mean squared error (MSE) loss,

$$L(Y_{t+h}, \hat{Y}_{t+h,t}) = (Y_{t+h} - \hat{Y}_{t+h,t})^2, \quad (1)$$

where  $L(\cdot)$  is the loss function,  $t$  is the time at which the forecast is computed,  $h$  is the forecast horizon,  $Y_{t+h}$  is the predicted variable, and  $\hat{Y}_{t+h,t}$  is the forecast based on information available at time  $t$ . Assuming covariance stationarity, under MSE loss, the optimal forecast and the associated forecast error satisfy a set of standard conditions (see, e.g., Diebold and Lopez 1996):

1. The optimal forecast of  $Y_{t+h}$  is its conditional expectation, so the forecast is conditionally (and unconditionally) unbiased.
2. The  $h$ -step-ahead forecast error exhibits zero serial covariance beyond lag  $(h - 1)$ .
3. The unconditional variance of the forecast error is a non-decreasing function of the forecast horizon.

Unfortunately, such properties often do not provide a useful guide to empirical tests because they do not generally hold under other loss functions (Patton and Timmermann 2006), and the forecaster's loss function is unknown in most applications. This is particularly important because there are often good reasons to believe that loss depends asymmetrically on positive and negative forecast errors, as has been discussed by, among others, Granger (1969), Varian (1974), Granger and Newbold

(1986), Zellner (1986), Weiss (1996), West, Edison, and Cho (1993), and Christoffersen and Diebold (1997).

As an important case in point, consider the one-quarter-ahead forecasts of real output [gross domestic product (GDP)] growth produced by the Board of Governors of the Federal Reserve—the so-called "Greenbook" forecasts—plotted in Figure 1 against realizations of annualized real GDP growth over the period 1968Q4–1999Q4, a total of 125 observations. The Federal Reserve does not explain how the Greenbook forecasts are constructed (although these forecasts are known to be based on judgmental information as well as on formal statistical models), nor does it publish the loss function used, either explicitly or implicitly, in constructing the forecasts. However, it is known that considerable resources are devoted to the production of these forecasts and that they are used by the Fed in the setting of monetary policy. Therefore, we should reasonably expect that these forecasts contain valuable information about future GDP growth. In fact, due to the sensitivity of these forecasts—and the information they contain about monetary policy—the Fed releases them only with a 5-year delay.

Figure 2 plots realized GDP growth against the forecast in a scatterplot. If the forecasts were optimal under MSE loss, then it follows from property 1 that they should fall around the 45-degree line. This property is commonly tested through a so-called Mincer–Zarnowitz regression of realized values on a constant and the forecast (Mincer and Zarnowitz 1969), the implication being that the associated coefficients should be 0 and 1. Ordinary least squares (OLS) parameter estimates (and standard errors) of this regression were

$$Y_{t+1} = 1.253 + .710 \hat{Y}_{t+1,t} + u_{t+1}$$

(.451)      (.106)

and

$$R^2 = .218.$$

Under MSE loss, forecast optimality does not rule out heteroscedasticity in the residuals from this regression. The only implication of optimality is that the forecast errors should follow a martingale difference sequence with respect to the information set used by the forecaster. Autocorrelation in the residuals,  $u_{t+1}$ , is ruled out, however, because lagged forecast errors are generally assumed to be part of the forecaster's information set. This implication, and other implications, of forecast optimality is not tested in the Mincer–Zarnowitz regression, which

Andrew J. Patton is Reader in Economics, University of Oxford, Oxford OX1 3UQ, United Kingdom (E-mail: [andrew.patton@economics.ox.ac.uk](mailto:andrew.patton@economics.ox.ac.uk)). Allan Timmermann is Professor of Economics and Management, University of California San Diego, La Jolla, CA 92093 (E-mail: [atimmerm@ucsd.edu](mailto:atimmerm@ucsd.edu)). Part of this work was done while Patton was Lecturer in Finance, London School of Economics. The authors thank three anonymous referees for very helpful comments and suggestions. Jeremy Berkowitz, Sean Campbell, Carlos Capistran, Peter Christoffersen, Valentina Corradi, Frank Diebold, Graham Elliott, Raffaella Giacomini, Clive Granger, Peter Hansen, Mark Machina, Nour Meddahi, Jim Powell, Barbara Rossi, Paul Ruud, Kevin Sheppard, Hal White, Stanley Zin, and seminar participants at Berkeley, Brown, CORE, Cornell, CREST, Erasmus University Rotterdam, LSE, McGill, Pennsylvania, Queen Mary, Stanford, UCLA, UC Riverside, UCSD, Warwick Business School, and the 2002 EC<sup>2</sup> conference all provided useful comments.

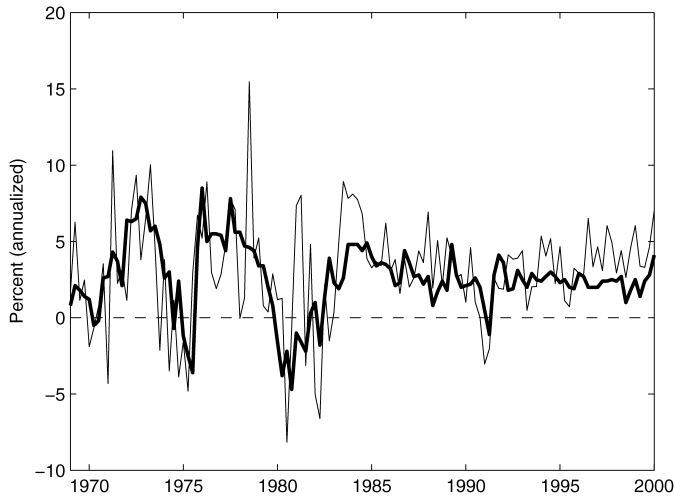


Figure 1. Real GDP growth (annualized) (—) and the Federal Reserve’s “Greenbook” forecasts of real GDP growth (—) over the period 1968Q4–1999Q4.

uses information only on the mean and correlation between the forecast,  $\hat{Y}_{t+1,t}$ , and the realization,  $Y_{t+1}$ , to test for unbiasedness. As such, this regression should be viewed as only a weak (necessary but not sufficient) test of one implication of forecast optimality under MSE loss. Even so, a Wald test of the joint restriction on the intercept and the slope yields a  $\chi^2_2$  statistic of 8.57 and a  $p$  value of .01. All standard errors in this article use the heteroscedasticity and autocorrelation robust estimator of Newey and West (1987). The test statistic and  $p$  value obtained when the usual OLS parameter covariance matrix is used are 8.90 and .01. In both cases, we reject the null of forecast optimality under MSE loss at the .05 level.

This strong rejection of optimality is surprising at first sight given the Fed’s incentives to forecast this variable well and the considerable resources devoted to its prediction. Keep in mind, however, that the rejection was predicated on MSE loss, whereas actually the Federal Reserve’s loss function is unknown. In fact, as pointed out by Nobay and Peel (2003) and

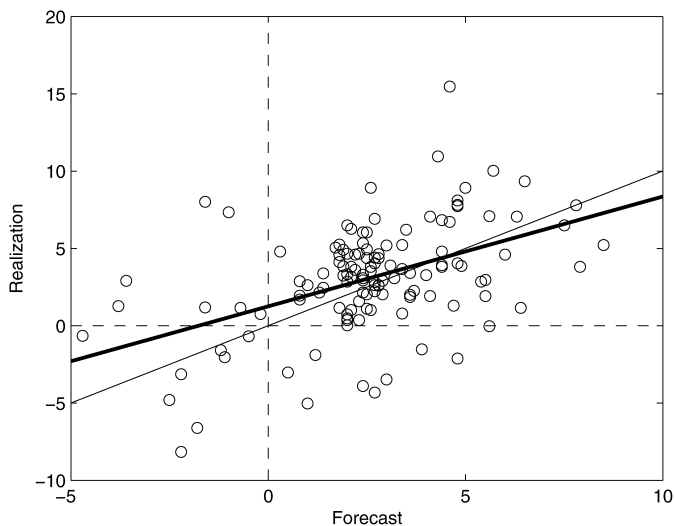


Figure 2. Realized real GDP growth (o) against forecast, with 45-degree line (—) and OLS fitted line (—).

Capistran (2006), there are strong reasons to assume that the Fed has asymmetric loss over forecast errors rather than MSE loss.

These empirical findings point to the need for establishing tests of forecast optimality that apply when the loss function is unknown. Toward this end, here we establish new results that trade-off restrictions on the forecaster’s loss function against restrictions on the data-generating process (DGP). For example, in situations where the conditional higher-order moments of the predicted variable are constant, we show that although the optimal forecast may well be biased, optimal  $h$ -step-ahead forecast errors display serial dependence of at most order  $h - 1$ , implying that optimal one-step-ahead forecasts generate serially uncorrelated forecast errors. These results hold irrespective of the shape of the loss function and offer new ways of testing forecast optimality that do not require knowledge of the specific loss function, but do require restrictions on the underlying DGP. They will be useful in the common situation where the shape of the loss function is unknown but the restrictions on the DGP can be tested empirically. We present similar results that hold when the DGP exhibits heteroscedasticity of a general unknown form and mild restrictions are imposed on the shape of the loss function, using a new family of quantile-based tests. We are not aware of any existing tests of forecast optimality that are robust to the loss function of the forecaster.

An important restriction required for some of these new results is that the loss function depends only on the forecast error. However, it is quite likely that the loss also depends on such factors as the level of the predicted variable or the value of some other state variable (cf. Garratt, Lee, Pesaran, and Shin 2003). For example, it may be particularly costly to overpredict GDP growth when growth is already very low. In effect, this would amount to the Fed signaling a false recovery, which could lead to an overly tight monetary policy at exactly the wrong point in time. To deal with broader classes of loss functions, we present a new method to test forecast optimality based on a flexible model of the loss function, which is then tested using a set of overidentifying restrictions. Indeed, empirical analysis suggests that this broader class of loss functions is needed to maintain the assumption of optimality for the Fed’s forecasts.

The outline of the article is as follows. Section 2 derives testable properties of optimal forecasts when the loss function is unknown but testable restrictions can be imposed on the DGP. Section 3 extends our results to the case where the loss function does not depend exclusively on the forecast error, Section 4 reports size and power results from a Monte Carlo simulation experiment, and Section 5 concludes. An Appendix presents technical details and proofs.

## 2. TESTABLE IMPLICATIONS UNDER UNKNOWN LOSS

Suppose that a decision maker is interested in forecasting some univariate time series,  $Y \equiv \{Y_t; t = 1, 2, \dots\}$ ,  $h$  steps ahead given information at time  $t$ ,  $\mathcal{F}_t$ . Let  $\tilde{\mathbf{Z}}_t$  be a  $(m \times 1)$  vector of predictor variables used by the forecaster (e.g., the Federal Reserve), and let  $\mathcal{F}_t$  be the  $\sigma$ -field generated by  $\{(Y_{t-k}, \tilde{\mathbf{Z}}'_{t-k})'; k \geq 0\}$ . We denote a generic element of  $\mathcal{F}_t$  as  $\mathbf{Z}_t$ , denote the conditional distribution of  $Y_{t+h}$  given  $\mathcal{F}_t$  as  $F_{t+h,t}$ ,

and denote the conditional density as  $f_{t+h,t}$ . Point forecasts conditional on  $\mathcal{F}_t$ , denoted by  $\hat{Y}_{t+h,t}$ , belong to  $\mathcal{Y}$ , a subset of  $\mathbb{R}$ , whereas forecast errors are given by  $e_{t+h,t} = Y_{t+h} - \hat{Y}_{t+h,t}$ . In general, the objective of the forecast is to minimize the expected value of some loss function,  $L(Y_{t+h}, \hat{Y}_{t+h,t})$ , which is a mapping from realizations and forecasts to the real line,  $L: \mathbb{R} \times \mathcal{Y} \rightarrow \mathbb{R}$ . An optimal forecast is thus defined as

$$\hat{Y}_{t+h,t}^* \equiv \arg \min_{\hat{y} \in \mathcal{Y}} E[L(Y_{t+h}, \hat{y}) | \mathcal{F}_t]. \tag{2}$$

We use  $E_t[\cdot]$  as shorthand notation for  $E[\cdot | \mathcal{F}_t]$ , the conditional expectation given  $\mathcal{F}_t$ . We also define the conditional variance,  $V_t[Y_{t+h}] = E[(Y_{t+h} - E[Y_{t+h} | \mathcal{F}_t])^2 | \mathcal{F}_t]$ , and the unconditional equivalents,  $E[\cdot]$  and  $V[\cdot]$ . Finally, we define  $\mu_{t+h,t} \equiv E_t[Y_{t+h}]$  and  $\sigma_{t+h,t}^2 \equiv V_t[Y_{t+h}]$ .

Establishing properties of optimal forecasts that may be tested when the loss function of the forecaster is unknown requires restrictions on the DGP. We consider DGPs with dynamics in the conditional mean and conditional variance, but no dynamics in the remainder of the conditional distribution. These classes of DGPs are quite broad and include autoregressive moving average (ARMA) processes and nonlinear regressions, possibly with dynamics in the conditional variance process.

### 2.1 Conditional Mean Dynamics Only

First consider the class of DGPs that satisfy the following conditional homoscedasticity condition:

*Assumption D1.* The DGP is such that  $Y_{t+h} = \mu_{t+h,t} + \varepsilon_{t+h}$ ,  $\varepsilon_{t+h} | \mathcal{F}_t \sim F_{\varepsilon,h}(0, \sigma_{\varepsilon,h}^2)$ , where  $F_{\varepsilon,h}(0, \sigma_{\varepsilon,h}^2)$  is some distribution with mean 0 and variance  $\sigma_{\varepsilon,h}^2$ , which may depend on  $h$  but does not depend on  $\mathcal{F}_t$ .

The restriction of dynamics only in the conditional mean implies that the innovation term,  $\varepsilon_{t+h}$ , is drawn from some distribution,  $F_{\varepsilon,h}$ , that generally will depend on the forecast horizon but is independent of  $\mathcal{F}_t$  and so is not denoted with a subscript  $t$ .

Following the extensive literature in economics and statistics, we concentrate initially on loss functions that satisfy the following assumption:

*Assumption L1.* The loss function depends solely on the forecast error, that is,  $L(y, \hat{y}) = L(y - \hat{y}) = L(e)$ ,  $\forall (y, \hat{y}) \in \mathbb{R} \times \mathcal{Y}$ .

Although Assumption L1 rules out certain loss functions, many common loss functions are of this form, including MSE, mean absolute error (MAE), lin–lin, and linear-exponential (linex) loss. Cases in which the loss depends not only on the size of the forecast error, but also on the level of the predicted variable are ruled out, however. As we show in Section 3, such cases arise in economics if the costs of prediction errors are not the same in recessions and expansions. They also can occur in meteorology, where underpredictions and overpredictions may be equally costly under normal weather conditions but may differ under more extreme conditions. (For example, underpredicting the strength of a hurricane or tornado could be particularly costly.)

Under these two assumptions, we obtain the following testable implications of forecast optimality.

*Proposition 1.* Suppose that the loss is a function solely of the forecast error (Assumption L1), whereas the DGP has dynamics only in the conditional mean (Assumption D1). Then the following results hold:

- a. The optimal forecast takes the form

$$\hat{Y}_{t+h,t}^* = \mu_{t+h,t} + \alpha_h^*, \tag{3}$$

where  $\alpha_h^*$  is a constant that depends on the distribution  $F_{\varepsilon,h}$  and the loss function.

- b. The forecast error associated with the optimal forecast,  $e_{t+h,t}^*$ , is independent of all  $\mathbf{Z}_t \in \mathcal{F}_t$ . In particular,  $\text{cov}(e_{t+h,t}^*, e_{t+h-j,t-j}^*) = 0$  for all  $j \geq h$  and any  $h > 0$ .

- c. The variance of the forecast error associated with the optimal forecast,  $V(e_{t+h,t}^*)$ , is nondecreasing in the forecast horizon,  $h$ .

All proofs are contained in the Appendix. This proposition shows that under a testable assumption on the DGP and only one weak assumption on the loss function, the forecast errors associated with the optimal forecast are serially uncorrelated at lags greater than or equal to the forecast horizon for any error-based loss function. Thus, given a sequence of realizations and forecasts,  $\{(Y_{t+h}, \hat{Y}_{t+h,t})\}_{t=1}^T$ , we may test for forecast optimality *without* knowledge of the forecaster’s loss function by testing, for example, the serial correlation properties of the forecast errors. For financial applications, the assumption of constant higher-order conditional moments is clearly too strong, but in many economic applications the assumption that all dynamics are driven by the conditional mean may be reasonable. In these cases, forecast optimality can be tested with a large degree of robustness to the loss function of the forecaster, for example, by testing  $\boldsymbol{\beta} = \mathbf{0}$  in such regressions as

$$e_{t+h,t} = \alpha + \boldsymbol{\beta}' \mathbf{Z}_t + u_{t+h}. \tag{4}$$

Note that no estimate of the conditional mean of  $Y_{t+h}$  is needed to conduct this test. This is important because  $\mu_{t+h,t}$  is generally unknown. This test exploits the result that all variables known to have been in the forecaster’s information set—except trivially for a constant—are orthogonal to the forecast error, which is observable without knowledge of  $\mu_{t+h,t}$ .

We focus on testing for correlation, although, of course, more generally we could test for complete independence between  $e_{t+h,t}$  and any  $\mathbf{Z}_t \in \mathcal{F}_t$ , or other implications of independence. For example, rather than using OLS to estimate the parameters in (4), we could use least absolute deviation or quantile regression (see Koenker and Bassett 1978).

The last part of the proposition reveals that optimality tests based on the variance of the forecast error being weakly increasing in the forecast horizon require restrictive assumptions on either the loss function (namely, that MSE loss applies) or the DGP (i.e., dynamics in the conditional mean only).

### 2.2 Conditional Mean and Variance Dynamics

We next provide results for a more general class of conditional scale-location processes that satisfy the following assumption:

*Assumption D2.* The DGP is such that  $Y_{t+h} = \mu_{t+h,t} + \sigma_{t+h,t} \eta_{t+h}$ ,  $\eta_{t+h} | \mathcal{F}_t \sim F_{\eta,h}(0, 1)$ , where  $F_{\eta,h}(0, 1)$  is some distribution with mean 0 and unit variance that may depend on  $h$ , but does not depend on  $\mathcal{F}_t$ .

This class of DGPs is very broad and includes most common volatility processes, including autoregressive conditional heteroscedasticity (ARCH) and stochastic volatility (see Engle 1982; Taylor 1982). It nests those of Assumption D1 at the cost of being more restrictive on the class of loss functions that we consider herein. Specifically, we assume that the loss function is homogeneous in the forecast error:

*Assumption L2.* The loss function is a homogeneous function solely of the forecast error.

This assumption implies that  $L(ae) = g(a)L(e)$  for some positive function  $g$ . Commonly used loss functions, such as MSE, MAE, lin–lin, and asymmetric quadratic loss, are all of this form. This requirement is needed to ensure that the units of the forecast (e.g., cents versus dollars) do not affect the optimal forecast beyond a scale adjustment. Although this holds for many loss functions encountered in economics and finance, the requirement does exclude such cases as linear-exponential loss.

With these two assumptions, we obtain the following testable implications of forecast optimality.

*Proposition 2.* Suppose that the loss function is homogeneous in the forecast error (Assumption L2), whereas the DGP can have conditional mean and variance dynamics (Assumption D2). Define the standardized forecast error associated with the optimal forecast as  $d_{t+h,t}^* = e_{t+h,t}^*/\sigma_{t+h,t}$ . Then the following results hold:

- a. The optimal forecast takes the form

$$\hat{Y}_{t+h,t}^* = \mu_{t+h,t} + \sigma_{t+h,t} \cdot \gamma_h^*, \quad (5)$$

where  $\gamma_h^*$  is a constant that depends only on the distribution  $F_{\eta,h}$  and the loss function.

- b.  $d_{t+h,t}^*$  is independent of any element  $\mathbf{Z}_t \in \mathcal{F}_t$ . In particular,  $\text{cov}(d_{t+h,t}^{*r}, d_{t+h-j,t-j}^{*s}) = 0$  for all  $j \geq h$  and any  $h > 0$  and all  $r$  and  $s$  for which the covariance exists.

Although the forecast error associated with the optimal forecast generally will not be unbiased, serially uncorrelated, or homoscedastic, the forecast error scaled by the conditional standard deviation will be independent of any  $\mathbf{Z}_t \in \mathcal{F}_t$ . This implies that  $d_{t+h,t}^*$  will be serially uncorrelated at lags  $j \geq h$  and homoscedastic. Therefore, forecast optimality could be tested by estimating nonparametric or flexible parametric models for the conditional mean and variance,

$$\begin{aligned} d_{t+h,t} &= \alpha_0 + g_1(\mathbf{Z}_t; \theta_1) + u_{t+h}, \\ u_{t+h} &= \sigma_{u,t+h} v_{t+h}, \quad v_{t+h} \sim (0, 1), \\ \sigma_{u,t+h}^2 &= \omega_0 + g_2(\mathbf{Z}_t; \theta_2). \end{aligned} \quad (6)$$

If  $g_1$  and  $g_2$  are defined such that  $g_1(\mathbf{z}; 0) = g_2(\mathbf{z}; 0) = 0$  for all  $\mathbf{z}$ , then a test of optimality may be obtained by testing  $H_0: \omega_0 = 1 \cap \theta_1 = \theta_2 = 0$  so that the conditional mean of  $d_{t+h,t}$  is time-invariant and its conditional variance equals unity.

The foregoing test is easily computed and does not require an estimate of  $\mu_{t+h,t}$ , although it does require that an estimate of  $\sigma_{t+h,t}^2$  be available. Under certain conditions, a consistent estimate of this conditional variance can be obtained from the observed  $Y_t$  process by either parametric methods [e.g., using

a generalized ARCH (GARCH)-type model] or nonparametric methods, using a realized volatility estimator (see, e.g., Andersen, Bollerslev, Diebold, and Labys 2001, 2003).

Researchers will not always have a reliable estimate of  $\sigma_{t+h,t}^2$  available, and so it would be particularly useful to establish under which conditions an optimality test can be based only on observables. Suppose that we restrict the first and second moment dynamics to be linked in a manner consistent with the widely used “constant coefficient of variation” model:

*Assumption D2’.* The DGP is such that  $Y_{t+h} = \beta \sigma_{t+h,t} + \sigma_{t+h,t} \eta_{t+h}$ ,  $\eta_{t+h} | \mathcal{F}_t \sim F_{\eta,h}(0, 1)$ , where  $\beta \in \mathbb{R}$  and  $F_{\eta,h}$  is some distribution with mean 0 and unit variance that may depend on  $h$  but does not depend on  $\mathcal{F}_t$ .

This “ARCH-in-mean” model is used extensively in financial applications, such as when the target variable is returns and expected returns are proportional to the level of risk as measured by the conditional standard deviation (see Engle, Lilien, and Robins 1987). Using this assumption, we obtain the following result.

*Corollary 1.* Suppose that the loss function is homogeneous in the forecast error (Assumption L2), whereas the DGP satisfies Assumption D2’ for  $\beta \neq -\gamma_h^*$ . Define  $\hat{d}_{t+h,t}^* \equiv (Y_{t+h} - \hat{Y}_{t+h,t}^*)/\hat{Y}_{t+h,t}$ . Then  $\hat{d}_{t+h,t}^*$  is independent of any element  $\mathbf{Z}_t \in \mathcal{F}_t$ . In particular,  $\text{cov}(\hat{d}_{t+h,t}^{*r}, \hat{d}_{t+h-j,t-j}^{*s}) = 0$  for all  $j \geq h$  and any  $h > 0$  and all  $r$  and  $s$  for which the covariance exists.

Note that the assumption that  $\beta \neq -\gamma_h^*$  is easily checked; if it is true, then the optimal forecast is identically zero for all  $t$ . Under the conditions of Corollary 1, we may test forecast optimality without specific knowledge of the loss function or any of the moments of the DGP, by testing that there is no serial correlation beyond lag  $h - 1$  in  $\hat{d}_{t+h,t} = (Y_{t+h} - \hat{Y}_{t+h,t})/\hat{Y}_{t+h,t}$ , and/or that the  $\hat{d}_{t+h,t}$  series is homoscedastic conditional on any  $\mathbf{Z}_t \in \mathcal{F}_t$ . This can be done simply through a regression of powers of  $\hat{d}_{t+h,t}$  on a constant and lags, of order greater than or equal to  $h$ , of various powers of  $\hat{d}_{t+h,t}$ , or as in (6).

### 2.3 Quantile Tests

Under the conditions of Propositions 1 or 2, it is possible to show that the optimal forecast can be expressed as a conditional quantile of the variable of interest. The usefulness of this result lies in the surprising finding that the optimal forecast is the *same* quantile at all points in time, though the quantile may change with the forecast horizon and is typically unknown, because it depends on the loss function. With such a representation, we can obtain an alternative test of forecast optimality using tests of quantile forecasts, eliminating the need to estimate the conditional variance of the variable of interest.

*Proposition 3.* Suppose that either (a) the loss is a function solely of the forecast error (Assumption L1), and the DGP has dynamics only in the conditional mean (Assumption D1), or (b) the loss function is homogeneous in the forecast error (Assumption L2) and the DGP has dynamics in the conditional mean and variance (Assumption D2). Then the following results hold:

a. The optimal forecast is such that, for all  $t$ ,

$$F_{t+h,t}(\hat{Y}_{t+h,t}^*) = q_h^*, \tag{7}$$

where  $q_h^* \in (0, 1)$  depends only on the distribution  $F_{\eta,h}$  and the loss function. If  $F_{t+h,t}$  is continuous and strictly increasing, then we can alternatively express this as

$$\hat{Y}_{t+h,t}^* = F_{t+h,t}^{-1}(q_h^*). \tag{8}$$

b. Let

$$I_{t+h,t}^* \equiv \mathbf{1}(Y_{t+h} \leq \hat{Y}_{t+h,t}^*),$$

where  $\mathbf{1}(A)$  equals 1 if  $A$  is true and 0 otherwise. Then  $I_{t+h,t}^*$  is independent of all  $\mathbf{Z}_t \in \mathcal{F}_t$ . In particular,  $I_{t+h,t}^* - q_h^*$  is a martingale difference sequence with respect to  $\mathcal{F}_t$ .

Note how assumptions on the loss function can be traded off against assumptions on the DGP. This proposition gives rise to a new test that is applicable even though  $q_h^*$  is unknown. The test simply projects the indicator function on elements in  $\mathcal{F}_t$  and an intercept and tests that  $\beta = \mathbf{0}$ ,

$$I_{t+h,t} = \alpha + \beta' \mathbf{Z}_t + u_{t+h}. \tag{9}$$

Alternatively, a logit model could be used to better reflect the binary nature of the dependent variable, or the likelihood ratio test of independence of Christoffersen (1998) could be used to test for serial dependence in  $I_{t+h,t}$ . If  $q_h^*$  is known, then it can be further tested that  $\alpha = q_h^*$ . Even in the common case where  $q_h^*$  is unknown, the key point to note is that the quantile test does not require knowledge of the true values of either  $\mu_{t+h,t}$  or  $\sigma_{t+h,t}$ . The test also does not require knowledge of the time-varying conditional distribution,  $F_{t+h,t}$ , which is unknown in practice. This can be compared with tests based on the probability integral transform of the data,  $F_{t+h,t}(Y_{t+h})$ , which should be uniform(0, 1) under the null that the forecasting model is specified correctly. Such tests can be difficult to implement given only a sequence of point forecasts, and without knowledge of the forecaster's loss function or information set, which generally is unobservable to the forecast evaluator. Due to this robustness property and the minimal information required for their implementation, we believe that this new class of quantile tests of forecast optimality is likely to find widespread use in empirical work.

Of the existing tests of forecast optimality in the literature, the one that allows for the greatest flexibility with respect to the loss function is due to Elliott, Komunjer, and Timmermann (2005). These authors derived tests of forecast optimality when the loss function is assumed to belong to the two-parameter family,

$$L(e_{t+1}; \alpha, p) \equiv [\alpha + (1 - 2\alpha)\mathbf{1}(e_{t+1} < 0)]|e_{t+1}|^p, \tag{10}$$

with a positive exponent  $p$  and an asymmetry parameter  $\alpha$ ,  $0 < \alpha < 1$ . This is very different from the approach proposed in this section, which does not constrain the loss function to belong to a prespecified parametric family of loss functions. In Section 3 we suggest an extension of the approach of Elliott et al. (2005) for applications where Assumption L1 or L2 does not hold.

Although Proposition 3 holds quite generally, it of course only applies to *point* forecasts. One possible extension of our work includes establishing results for density forecasting, an

area recently considered by Garratt et al. (2003) and Campbell and Diebold (2005), although applications of such results to empirical data will be limited, because currently the vast majority of published forecasts are point forecasts.

### 2.4 Nonhomogeneous Loss or Dynamics in Higher-Order Moments

The foregoing results under unknown loss constitute an exhaustive set of testable properties in the following sense. First, suppose that the loss function is not homogeneous (Assumption L2). Allowing for conditional variance dynamics (as in Assumption D2) then makes it difficult to obtain testable implications of forecast optimality that are robust to the loss function. Further, when there are dynamics in third-order or higher-order moments and Assumption D2 fails to hold, it is generally difficult to obtain easily tested results even if homogeneity is imposed on the loss function. To see this, consider the following, more general DGP:

*Assumption D3.* The DGP is such that  $Y_{t+h} = \mu_{t+h,t} + \sigma_{t+h,t}\eta_{t+h}$ ,  $\eta_{t+h}|\mathcal{F}_t \sim F_{\eta,t+h,t}(0, 1)$ , where  $F_{\eta,t+h,t}(0, 1)$  is some time-varying distribution, with mean 0 and unit variance, that depends on  $\mathcal{F}_t$  and possibly also on  $h$ .

This class of DGPs nests those of Assumption D2, because we allow for a time-varying conditional mean, conditional variance, and other properties of the distribution (e.g., time-varying conditional skew or kurtosis). We obtain the following ‘‘impossibility’’ result.

*Proposition 4.* a. Suppose that the DGP has conditional mean and variance dynamics (Assumption D2), the loss is solely a function of the forecast error (Assumption L1), but that the loss function is not homogeneous in the forecast error (Assumption L2 is violated). Then

$$\hat{Y}_{t+h,t}^* = \mu_{t+h,t} + \alpha_{t+h,t}^*, \tag{11}$$

where  $\alpha_{t+h,t}^*$  is a scalar that depends on the loss function,  $\sigma_{t+h,t}$ , and  $F_{\eta,h}$ .

b. Suppose the loss function is homogeneous in the forecast error (Assumption L2) and that the DGP has dynamics beyond the conditional mean and variance (Assumption D3). Then

$$\hat{Y}_{t+h,t}^* = \mu_{t+h,t} + \sigma_{t+h,t} \cdot \gamma_{t+h,t}^*, \tag{12}$$

where  $\gamma_{t+h,t}^*$  is a scalar that depends on the loss function and  $F_{\eta,t+h,t}$ .

Under the conditions for part a or b of the foregoing proposition, the forecast error associated with the optimal loss,  $e_{t+h,t}^* \equiv Y_{t+h} - \hat{Y}_{t+h,t}^*$ ; its standardized equivalent,  $d_{t+h,t}^* = e_{t+h,t}^*/\sigma_{t+h,t}$ ; and the indicator variable,  $I_{t+h,t}^* \equiv \mathbf{1}(Y_{t+h} \leq \hat{Y}_{t+h,t}^*)$ , will all generally be correlated with some  $\mathbf{Z}_t \in \mathcal{F}_t$  in a way that depends on the unknown loss function. Thus if the loss function is not homogeneous, or if there are higher-order dynamics in the DGP, then these objects generally will not be useful for testing forecast optimality. In these cases it is difficult to obtain testable restrictions on the forecast error that do not require knowledge of the shape of the loss function, even if  $\sigma_{t+h,t}$  is known (or, more generally, even if  $F_{t+h,t}$  is known).

Of course, if the loss function and the DGP are known, optimality properties of the forecast can be derived directly. For example, assuming linex loss and a first-order Markov Gaussian mixture model, Patton and Timmermann (2006) derived the properties of the optimal forecast and forecast errors.

### 2.5 Empirical Results

We next return to the Federal Reserve Greenbook forecasts of output growth to demonstrate the theoretical results. These are the data shown in Figure 1. These forecasts are all one-quarter-ahead forecasts, so  $h = 1$ .

Our realized values of GDP are taken as the 2006Q1 “vintage” of the real GDP growth figures over the sample period, obtained from the Federal Reserve Bank of Philadelphia’s web page. This data were studied in depth by Croushore and Stark (2001). The data and a detailed description of its construction are available at <http://www.phil.frb.org/econ/forecast/reaindex.html>. In Section 2.5.1 we discuss the results obtained using an earlier vintage of real GDP growth data and show that they are similar to those obtained using the most recent vintage.

Under the joint assumption that the Fed’s forecasts are optimal, its loss is solely a function of the forecast error (Assumption L1), and there are no dynamics beyond the conditional mean (Assumption D1) of real GDP growth, Proposition 1 shows that the Greenbook forecast errors would be conditionally homoscedastic. We test for conditional heteroscedasticity in the Greenbook forecast errors using Engle’s (1982) test, which is based on the null that  $\alpha_i = 0$  for all  $i \geq 1$  in the regression,

$$e_{t+1,t}^2 = \alpha_0 + \alpha_1 e_{t,t-1}^2 + \dots + \alpha_L e_{t-L+1,t-L}^2 + u_{t+1}.$$

A test with four lags generated a  $p$  value of .03, indicating the presence of conditional heteroscedasticity. This suggests either that the Fed’s forecasts are suboptimal or that Assumption D1 or L1 is violated.

Testing Assumption D1 directly requires modeling the conditional mean of output growth. We used an ARMA(1, 1) model for the conditional mean, which was found to remove all significant serial correlation, and then tested for serial correlation in the squared residuals from this model. The  $p$  values from this test were .65 for lag 1, .09 for lag 4, and .01 for lag 8. These results indicate the presence of conditional heteroscedasticity and hence suggest that Assumption D1 is an inappropriate assumption for quarterly output growth.

The previous results suggest that we need to allow for more general dynamics in the DGP for real GDP growth. For this purpose, we tested Assumption D2 directly on the real GDP growth series by estimating an ARMA(1, 1) model for the conditional mean and a GARCH(1, 1) model for the conditional variance (cf. Bollerslev 1986). Simple, parsimonious GARCH models have been shown to work well in numerous other studies of macroeconomic and financial time series. We tested for serial correlation in the first four powers of the standardized residuals and found no evidence of any serial correlation out to eight lags, suggesting that Assumption D2 may be an appropriate assumption for this data. We also could have used the generalized tests of Hong (1999) to test Assumption D2.

To test the joint hypothesis that forecasts are optimal, the loss function is homogeneous in the forecast error, and the

Table 1. Summary statistics

	Realization	Forecast	Forecast error
Mean	3.10	2.59	.50
Standard deviation	3.54	2.33	3.20
Skewness	-.22	-.41	.08
Kurtosis	4.46	4.03	3.87
Minimum	-8.16	-4.70	-7.01
Maximum	15.46	8.50	10.86
Autocorrelation 1	.26*	.73*	.01
Autocorrelation 2	.20*	.55*	.10
Autocorrelation 3	.08	.45*	-.02
Autocorrelation 4	.05	.34*	.07
Jarque-Bera statistic	12.04*	9.06*	3.53
Jarque-Bera $p$ value	.002	.011	.171

NOTE: This table presents basic summary statistics on the realized real GDP growth, the Federal Reserve’s “Greenbook” forecasts of GDP growth, and the corresponding forecast errors over the period 1968Q4–1999Q4. The rows labeled “Autocorrelation  $j$ ” report the  $j$ -th-order autocorrelation of the series. Autocorrelations that are significant at the .05 level are marked with an asterisk. The last two rows present the Jarque-Bera test statistic and  $p$  value (Jarque and Bera 1987), testing the null hypothesis that the forecast errors are normally distributed.

more general Assumption D2 (which allows for conditional heteroscedasticity) holds for our GDP data, we use the result from Proposition 2 that under these assumptions, the forecast error associated with the optimal forecast will be of the following form:

$$e_{t+1,t} = \gamma_1 \sigma_{t+1,t} + \sigma_{t+1,t} \eta_{t+1}, \tag{13}$$

$$\eta_{t+1,t} | \mathcal{F}_t \sim F_{\eta,1}(0, 1).$$

Under the assumption of Corollary 1, the forecast  $\hat{Y}_{t+1,t}$  already incorporates information on the conditional mean of  $Y_{t+1}$ , and so an independent estimate of  $\mu_{t+1,t}$  is not required in this test. If we knew  $\sigma_{t+1,t}$  then we could construct  $d_{t+1,t} \equiv e_{t+1,t} / \sigma_{t+1,t}$ , which, under the conditions for Proposition 2, will be serially uncorrelated, be conditionally homoscedastic, and have constant conditional higher-order moments. To implement a test, we model the conditional variance as a simple GARCH(1, 1) process, allowing for the GARCH-in-mean effects implied by Proposition 2. Of course, the possibility remains that this model is misspecified and that this affects our conclusions. The GARCH-in-mean parameter,  $\gamma_1$ , was significant (with a  $p$  value  $< .01$ ), presenting further evidence against the optimality of these forecasts under MSE loss. Simple tests for serial correlation in the third and fourth powers of the forecast errors standardized by the GARCH estimates of conditional variance revealed no evidence against Assumption D2, conditional on the forecasts being optimal and Assumption L2 being satisfied.

If we could further assume that Assumption D2’ holds for real GDP growth, then we could apply Corollary 1 and use the Greenbook forecasts to standardize the forecast errors. Under the joint assumption that Assumptions D2’ and L2 hold and forecasts are optimal,  $\hat{d}_{t+1,t} = (Y_{t+1} - \hat{Y}_{t+1,t}) / \hat{Y}_{t+1,t}$  should be independent of any element in the forecaster’s information set, including past (standardized) forecast errors and any transformation of these. Therefore, we tested for serial correlation in  $\hat{d}_{t+1,t}$  and  $\hat{d}_{t+1,t}^2$ . We found no evidence of serial correlation in

$\hat{d}_{t+1,t}$  at any lag up to 8, but found significant evidence of serial correlation in  $\hat{d}_{t+1,t}^2$ . The  $p$  values from a Ljung–Box test with four lags was .01, indicating significant serial correlation. This indicates that the joint hypothesis that Assumption D2' holds, loss is homogeneous in the forecast error (Assumption L2), and the Greenbook forecasts are optimal can be rejected.

Given the plausibility of Assumption D2 for real GDP growth, and the violation of Assumption D2', we now use the test of forecast optimality based on Proposition 3. Rather than rely on a (possibly misspecified) parametric estimator of the conditional variance of the real GDP growth series, we implement the test of forecast optimality under unknown loss using the indicator variable  $I_{t+1,t} \equiv \mathbf{1}(Y_{t+1} \leq \hat{Y}_{t+1,t})$ . This variable should be independent of all elements of  $\mathcal{F}_t$  under the null of forecast optimality and requires no specification of the mean or variance. We consider two parsimonious tests of this restriction, both obtained via OLS regression,

$$I_{t+1,t} = .346 + .036 \hat{Y}_{t+1,t} + u_{t+1}$$

(.071)    (.016)

and

$$I_{t+1,t} = .334 + .039 \hat{Y}_{t+1,t} + .036 I_{t,t-1} + u_{t+1}$$

(.095)    (.015)    (.108)

The first regression reveals significant evidence of correlation between the indicator variable and the forecast; the  $t$  statistic is 2.27, which is significant at the .05 level. In the second regression, when we include both the lagged indicator variable and the forecast, we find that the coefficient on the lagged indicator is not significantly different from 0, whereas the coefficient on the forecast is significant at the .05 level. A joint test that both coefficients are equal to 0 yields a  $p$  value of .06, meaning that the null is not rejected at the .05 level but is rejected at the .10 level.

Because our tests revealed no evidence against Assumption D2, these results constitute evidence against the optimality of the Greenbook forecasts under *any* loss function that is homogeneous in the forecast error. But if the Fed's loss function does not satisfy this restriction, then these forecasts may yet be optimal under a more general loss function. Overall, we take these regressions and the earlier results as evidence against the optimality of the Greenbook forecasts under unknown loss functions that are homogeneous in the forecast error.

**2.5.1 Real-Time Data.** In addition to the 2006Q1 “vintage” of real GDP growth figures, we also consider using “real-time” GDP figures, that is, the figures that historically would have been available to forecasters at each point in time. Diebold and Rudebusch (1991) and Croushore and Stark (2001) discussed the differences between using revised data versus real-time data. These data take into account the possible effect of measurement errors on the forecasting performance as measured historically and use the fact that the Fed may have as its forecast target the initial release of GDP figures, rather than the underlying true GDP figures, as proxied by the latest vintage data. Following studies such as those of Romer and Romer (2000) and Capistran (2006), we use the second revision of the real GDP growth figures.

The results that we obtain are consistent with those reported earlier for the final revision. For example, we find evidence

against optimality under MSE loss using these data from the following regression:

$$e_{t+1,t} = 1.046 - .300 \hat{Y}_{t+1,t} + .072 e_{t,t-1} + u_{t+1}$$

(.985)    (.253)    (.111)

Although none of the coefficients in this model are individually significant at the .05 level, a joint test that all parameters equal 0 yields a  $\chi^2_3$  statistic ( $p$  value) of 8.56 (.04), implying rejection of this restriction at the .05 level. We also find significant serial correlation in the real-time forecast errors, indicating either that Assumptions D1 and L1 do not hold or that the forecasts are suboptimal. Furthermore, using the real-time data, the indicator variable regressions yielded

$$I_{t+1,t} = .362 + .027 \hat{Y}_{t+1,t} + u_{t+1}$$

(.067)    (.019)

and

$$I_{t+1,t} = .255 + .028 \hat{Y}_{t+1,t} + .250 I_{t,t-1} + u_{t+1}$$

(.076)    (.019)    (.087)

Thus for the real-time data, the coefficient on the forecast is not significant, whereas the coefficient on the lagged indicator is now highly significant. The  $\chi^2_2$  statistic and  $p$  value from the second regression are 10.07 and .01, and so again we strongly reject optimality in conjunction with Assumptions D2 and L2.

**2.5.2 The “Great Moderation.”** A factor that may affect our results is the presence of a structural break in the variance of U.S. GDP growth, known as the “great moderation,” generally considered to have occurred around 1984Q1 (see McConnell and Perez-Quiros 2000; Stock and Watson 2002). Breaks such as these are widely considered a key source of forecast failure (cf. Clements and Hendry 1998, 2006). To explore this possibility, we carried out separate tests using data up to 1983Q4 (61 observations) and from 1984Q1 onward (64 observations). The subsample results were very similar to those obtained from the full sample; Mincer–Zarnowitz tests of forecast optimality under MSE loss yielded the following results:

$$\text{Pre-1984 sample: } Y_{t+1} = .953 + .692 \hat{Y}_{t+1,t} + u_{t+1}$$

(.655)    (.118)

$$R^2 = .225$$

and

$$\text{Post-1984 sample: } Y_{t+1} = 1.347 + .782 \hat{Y}_{t+1,t} + u_{t+1}$$

(1.127)    (.381)

$$R^2 = .144$$

The  $\chi^2_2$  statistics ( $p$  values) that the intercepts equal 0 and the slope coefficients equal 1 were 7.19 (.03) in the first subsample and 7.66 (.02) in the second subsample. Thus optimality under MSE loss is rejected in both subsamples, as well as in the full sample.

The quantile-based test for optimality under unknown, homogeneous loss yielded the following results for the two subsamples:

Pre-1984 sample:

$$I_{t+1,t} = .556 + .025 \hat{Y}_{t+1,t} - .246 I_{t,t-1} + u_{t+1}$$

(.108)    (.020)    (.127)

and

Post-1984 sample:

$$I_{t+1,t} = \underset{(.171)}{.059} + \underset{(.058)}{.085} \hat{Y}_{t+1,t} + \underset{(.121)}{.288} I_{t,t-1} + u_{t+1}.$$

The  $\chi^2_2$  statistics ( $p$  values) that the both slope coefficients equal 0 in each regression were 6.09 (.05) and 7.41 (.02) in the first and second subsamples. Again, forecast optimality in conjunction with Assumptions L2 and D2 is rejected in both subsamples, as in the full sample. The similarity between the results in the two subsamples and the results from the full data sample, under both MSE and unknown loss, provides some assurance that a structural break in U.S. GDP growth volatility is not driving our findings.

### 3. TESTABLE IMPLICATIONS UNDER GENERAL DATA-GENERATING PROCESSES

The rejection of optimality for the Greenbook forecasts reported in the previous section is surprising; the null hypothesis is very general and covers a large class of loss functions. One might expect that allowing for such a wide range of different loss functions would erode the power of our tests, particularly when applied to a sequence of forecasts that we expect *ex ante* to be “good,” such as those from the Federal Reserve. As shown in our simulation study, presented in Section 4, this turns out to not be the case.

However, it might realistically be the case that the Federal Reserve’s loss function cannot be assumed to be solely a function of the forecast error. To a conservative policy maker, overpredictions of economic growth are likely not only to be more costly than underpredictions but also to be disproportionately more costly in periods of low growth, because it may incorrectly signal a recovery from a recession. This in turn is likely to lead to wrong choices in how monetary policy is set. This points to a need to consider forecast optimality in situations where not only the forecast error, but also the level of the predicted variable matter. For such cases, it is possible to construct a test based on a flexible parametric estimate of the first derivative of the loss function with respect to  $\hat{y}$ . The first-order condition  $E_t[\partial L(Y_{t+h}, \hat{Y}_{t+h,t}^*)/\partial \hat{y}] = 0$  implies that  $E[\partial L(Y_{t+h}, \hat{Y}_{t+h,t}^*)/\partial \hat{y} \cdot \mathbf{Z}_t] = 0$  for any  $\mathbf{Z}_t \in \mathcal{F}_t$ . For notational simplicity, let

$$\lambda(y, \hat{y}) \equiv \frac{\partial L(y, \hat{y})}{\partial \hat{y}}. \tag{14}$$

For example, we may obtain a flexible parametric estimate of  $\lambda(y, \hat{y})$ , denoted by  $\lambda(y, \hat{y}; \theta)$ , based on a linear spline model. To see how a linear spline could be used to approximate the function  $\lambda(y, \hat{y})$ , assume initially that  $\lambda = \partial L(e)/\partial e$ , let  $(\zeta_1, \dots, \zeta_K)$  be the nodes of the spline, and impose that one of the nodes is 0. We impose that the spline is continuous, although not necessarily differentiable, except possibly at 0. We could allow discontinuities in  $\lambda$  at the cost of introducing more parameters to estimate.

With just a few nodes, this class of loss functions is very flexible, nesting both MSE and MAE as special cases, as well as lin–lin, the symmetric nonconvex loss function of Granger (1969), and the class of loss functions used by Elliott et al. (2005). If we further impose that the spline is continuous at 0,

then MSE loss is nested in the interior of the parameter space, at the cost of the MAE and lin–lin loss functions not being nested. In this case the resulting estimated loss function is a piecewise quadratic spline with piecewise linear derivative,  $\lambda(e; \theta)$ , which is continuous and differentiable everywhere (except at the  $K$  nodes),

$$\frac{\partial \lambda(e; \theta)}{\partial e} = \begin{cases} \gamma_1, & \text{for } e \leq \zeta_1 \\ \gamma_i, & \text{for } \zeta_{i-1} < e \leq \zeta_i, i = 2, \dots, K \\ \gamma_{K+1}, & \text{for } e > \zeta_K, \end{cases} \tag{15}$$

where  $\theta = [\gamma_1, \gamma_2, \dots, \gamma_{K+1}]'$ . Here  $\lambda(e; \theta)$  and  $L(e; \theta)$  are constructed from the foregoing specification by imposing that  $\lambda(0; \theta) = L(0; \theta) = 0$  and that both  $\lambda(e; \theta)$  and  $L(e; \theta)$  are continuous in  $e$ . Because  $\lambda(e; \theta)$  is identified only up to a multiplicative constant, some normalization is needed to identify the parameters; for example, we could impose that  $\sum_{i=1}^{K+1} \gamma_i = 1$ . Furthermore, it is important to impose constraints on  $\theta$  so that the resulting estimate of  $\lambda$  satisfies the assumptions required for it to be the first derivative of some valid loss function, for example, that the loss function is weakly increasing in the absolute value of the forecast error.

In applications where we have reason to assume that the loss from a forecast is solely a function of the forecast error (i.e., Assumption L1 is satisfied), the problem simplifies to approximating the function  $\lambda(y - \hat{y}) = \lambda(e)$ . In other applications, such a restriction may not be well-founded, and so no such simplification is available. In this case we must use a more flexible specification to approximate the function  $\lambda(y, \hat{y})$  or, equivalently,  $\lambda(e, y)$ . Treating  $\lambda(e, y)$  rather than  $\lambda(y, \hat{y})$  makes it simpler to impose the required conditions on  $\lambda$ . We propose the following specification, which is structured so that MSE loss is obtained when all parameters are set equal to 0:

$$\frac{\partial \lambda(e, y; \theta)}{\partial e} = \begin{cases} \gamma_1(y) \equiv \Gamma(\varphi_{01} + \varphi_{11}y - \ln K) & \text{for } e \leq \zeta_1 \\ \gamma_i(y) \equiv \left(1 - \sum_{j=1}^{i-1} \gamma_j\right) \cdot \Gamma(\varphi_{0i} + \varphi_{1i}y - \ln K) & \text{for } \zeta_{i-1} < e \leq \zeta_i, i = 2, \dots, K \\ \gamma_{K+1}(y) = 1 - \sum_{j=1}^K \gamma_j & \text{for } e > \zeta_K, \end{cases} \tag{16}$$

where  $\Gamma(x) \equiv (1 + \exp(-x))^{-1}$  is the logistic transformation. This specification allows  $y$  to affect the slopes of  $\lambda$ , guarantees that all slopes are weakly positive, and that the sum of the slopes equals 1. Of course, alternative ways of restricting the  $\gamma_i$ ’s to be nonnegative are possible by, for example, using a probit instead of the logistic function or by including additional powers of  $y$  at the cost of having to estimate more parameters. These issues can be addressed in empirical work by means of a sensitivity analysis with regard to the assumed form of  $\partial \lambda/\partial e$ .

Under standard regularity conditions, the parameter vector of the approximating function can be estimated through the gener-



alized method of moments (GMM),

$$\hat{\theta}_T \equiv \arg \min_{\theta \in \Theta} \mathbf{g}_T(\theta)' \mathbf{W} \mathbf{g}_T(\theta), \tag{17}$$

$$\mathbf{g}_T(\theta) \equiv \frac{1}{T} \sum_{t=1}^T \lambda(e_{t+h,t}, \hat{Y}_{t+h,t}; \theta) \cdot \mathbf{Z}_t,$$

where  $\mathbf{W}$  is a weighting matrix and  $\Theta$  is a compact set. A test of forecast optimality can be obtained from a test of overidentifying restrictions if we ensure that we have more moment restrictions,  $k$ , than parameters,  $p$ ,

$$T \mathbf{g}_T(\hat{\theta}_T)' \hat{\mathbf{W}}_T^* \mathbf{g}_T(\hat{\theta}_T) \Rightarrow \chi_{k-p}^2, \text{ as } T \rightarrow \infty. \tag{18}$$

Here  $\hat{\mathbf{W}}_T^*$  is a consistent estimate of the optimal weight matrix (cf. Newey and McFadden 1994). This test of forecast optimality does not rely on any restrictions on the DGP other than standard conditions required for GMM estimators to be consistent and asymptotically normal. It does, however, rely on the linear spline being sufficiently flexible to approximate the unknown loss function. Thus a rejection of forecast optimality may be due either to a true failure of forecast optimality or to a failure of the approximation of the forecaster's loss function.

In contrast to the analysis of Elliott et al. (2005), our tests do not assume that the loss function belongs to a two-parameter family, nor do we need to restrict the loss function to depend only on the forecast error. Indeed, when the latter restriction holds, we recommend using the quantile-based test presented in Proposition 3.

### 3.1 Empirical Results

We now apply the spline-based tests of forecast optimality to the Greenbook forecasts. As outlined earlier, we use a quadratic spline for the loss function. Initially, we assume that the loss function is a function solely of the forecast error. We use three nodes,  $[-2, 0, 2]$ , which correspond to the .17, .44, and .70 quantiles of the empirical distribution of the forecast errors. With just three nodes, we require that  $\gamma_i \geq 0$  for all  $i$  for  $\lambda$  to correspond to the first derivative of a valid loss function. When the number of nodes exceeds three, nonnegative loss and continuity of the derivatives does not rule out that some of the  $\gamma_i$  values associated with the middle segments are negative. However, the central (adjacent to 0) and outermost segments must have nonnegative  $\gamma_i$  values so as not to violate these restrictions. In the estimation procedure, we normalize the function by imposing  $\sum_{i=1}^4 \gamma_i = 1$ .

As instruments for the moment conditions, we use a constant, the contemporaneous value of the forecast, and one lag each of the forecast error, realized GDP growth, and the loss function derivative,  $\lambda(y, \hat{y})$ . Thus we have five moment conditions and three free parameters. The estimated loss function is presented in Figure 3. This figure reveals that the estimated loss function is asymmetric, penalizing negative forecast errors (overpredictions) more than positive forecast errors (underpredictions). The average ratio of the loss from a negative forecast error to a positive error of the same magnitude [i.e.,  $L(-e)/L(e)$ ] for errors in the empirically relevant range  $[0, 10]$  is 1.44, with minimum and maximum values of .52 and 2.76.

The test of forecast optimality under the estimated loss function, obtained by using the two overidentifying moment con-

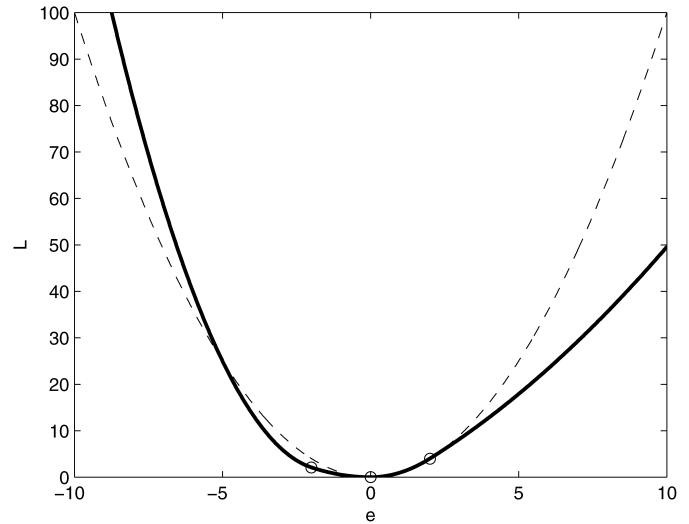


Figure 3. Estimated loss function of the Federal Reserve for real GDP growth forecasts, based on a quadratic spline with nodes  $[-2, 0, 2]$ , imposing that the loss function is solely a function of the forecast error (—), and MSE loss (---).

ditions, yields a  $\chi^2$  statistic ( $p$  value) of 5.57 (.06). A rejection of optimality using this test is consistent with the results in the previous section, which held for all loss functions that are homogeneous in the forecast error. The spline-based loss functions considered here are generally not homogeneous, and so they relax that restriction, but they are still constrained to be functions solely of the forecast error. Nevertheless, the empirical results are conditional on our choice of a specific number of nodes (three) used in the spline function.

To test the sensitivity of these results to the assumption of a logit specification as in (16), we also implemented a probit-type specification and obtained very similar results. For example, the  $p$  value from the  $J$  test for the probit specification of the spline model was .05, compared with .06 for the logit specification.

Finally, we estimate a more flexible specification of the Fed's loss function, allowing the loss function to depend on the forecast and the realization separately, rather than solely through the forecast error. As discussed earlier, we model  $\lambda$  as a function of  $(e, y)$  rather than  $(y, \hat{y})$ , because it is simpler to impose the required constraints on the former than on the latter. With three nodes, this model has a total of six free parameters, namely the intercept and slope parameters  $(\varphi_{01}, \varphi_{02}, \varphi_{03}, \varphi_{11}, \varphi_{12}, \varphi_{13})$  in the logistic specification in (16), compared with three in the simpler case  $(\varphi_{01}, \varphi_{02}, \varphi_{03})$ . We used as additional instruments one extra lag of the forecast error, realized GDP growth, and the derivative  $\lambda(e, \hat{y})$ . We plot the estimated loss function in Figure 4. To show the impact of realized GDP growth on the loss, we plot the loss as a function of the forecast error when realized GDP growth is fixed at its unconditional .25, .5, and .75 quantiles, corresponding to periods of low, average, or high economic growth. This figure shows that the level of realized GDP growth has substantial impact on the degree of asymmetry in the loss function. During periods of high economic growth, the loss function is approximately symmetric, whereas when realized GDP growth is at a lower level, the asymmetry becomes more pronounced, and, as in the simpler model, overpredictions are penalized more heavily than underpredictions. The average

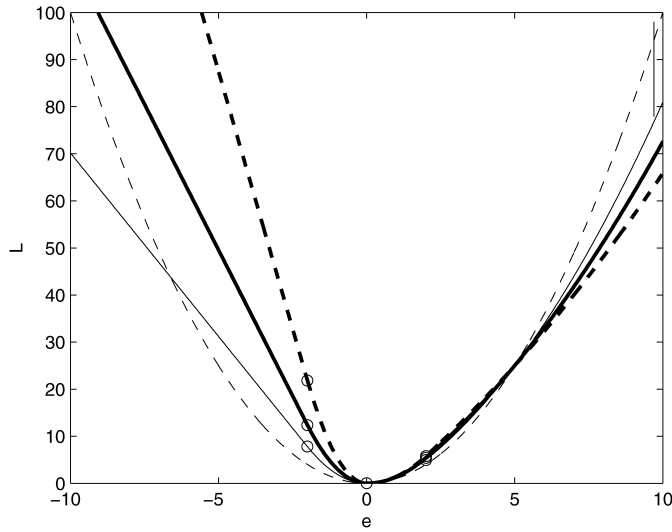


Figure 4. Estimated loss function of the Federal Reserve for real GDP growth forecasts, based on quadratic splines with nodes  $[-2, 0, 2]$ . This model allows the level of GDP growth to also affect the loss function; the estimated loss function is evaluated for GDP growth equal to its .25, .5, and .75 quantiles (---,  $Qy = .25$ ; —,  $Qy = .5$ ; —,  $Qy = .75$ ) and MSE loss (---).

ratio of the loss from a negative forecast error to a positive error of the same magnitude [i.e.,  $L(-e, y)/L(e, y)$ ] for errors in the range  $[0, 10]$  and for real GDP growth equal to its .25, .5, and .75 quantiles is 3.48, 1.97, and 1.26. Thus negative errors (overpredictions) are more than three times as costly to the Fed as positive errors (underpredictions) when GDP growth is low, whereas negative errors are only about 25% more costly than positive errors when GDP growth is high.

Using the two overidentifying moment conditions to test the optimality of the Greenbook forecasts based on this more flexible model yields a  $\chi^2$  statistic ( $p$  value) of .02 (.99). To check the robustness of this result, we extended the flexible spline model to allow not only  $y$ , but also  $y^2$  to affect the slopes of  $\lambda$ . We implemented this model using an extra four lags of the forecast errors as additional moment conditions, and found an estimated loss function very similar to those shown in Figure 4. The  $p$  value from the  $J$  test of the overidentifying restrictions for this more flexible model was .56, compared with .99 for the benchmark case.

These large  $p$  values imply that we have no evidence against the optimality of the Fed’s real GDP forecasts under this flexible loss function. This finding suggests that adequate modeling of the Fed’s loss function requires allowing both the forecast and the realization of real GDP growth to enter separately into the loss function, rather than just through their difference. This is an important result for future researchers studying the Fed’s forecast performance and objectives.

Our most flexible estimated loss function is consistent with a scenario in which the Fed seeks to issue “conservative” estimates of future economic growth; underestimates are penalized less heavily than overestimates. This conservatism is particularly important when economic growth is moderate or low. Issuing conservative forecasts means that actual economic growth comes in above the forecast more often than not (actual GDP growth was greater than the predicted value in 56% of the quarters in our sample), leading to a positive “surprise.”

Sorting on the predictions,  $\hat{y}$ , we found that for the lowest 25% of predicted values, there was a 65% probability that the actual value is higher than the predicted value. In comparison, for the top 75% of predicted values, only 45% of the actual values are above the predicted values. This is again consistent with the forecaster attempting to avoid overpredicting output growth in low-growth states of the world (as reflected in a current low estimate of  $\hat{y}$ ).

#### 4. SIMULATION RESULTS

To shed light on the finite-sample properties of the tests considered so far, we present the results of a small simulation study tailored to capture properties of our data. Spline models of the loss function are new to the literature, and GMM estimation is known to sometimes have problems in finite samples (see Hall 2005); thus a study of these tests in finite samples is of potential value. We use a simple but representative AR(1)–GARCH(1, 1) model as the DGP for the simulation,

$$\begin{aligned}
 Y_t &= .5Y_{t-1} + \sigma_t \varepsilon_t, & t = 1, 2, \dots, T; \\
 \sigma_t^2 &= .1 + .8\sigma_{t-1}^2 + .1\sigma_{t-1}^2 \varepsilon_{t-1}^2; \\
 \varepsilon_t &\sim \text{iidN}(0, 1).
 \end{aligned}
 \tag{19}$$

We consider three sample sizes ( $T = 100, 250, 1,000$ ) and two loss functions, MSE loss and an asymmetric quadratic loss function,

$$L(e; a) = \begin{cases} ae^2, & e > 0 \\ e^2, & e \leq 0, \end{cases}
 \tag{20}$$

where we set  $a = 1.84$ . Both of these loss functions are homogeneous in the forecast error, and so, by Proposition 2, we know that the optimal one-step-ahead forecast will take the form

$$\hat{Y}_{t+1,t}^* = \mu_{t+1,t} + \sigma_{t+1,t} \gamma_1^*.
 \tag{21}$$

Under MSE loss, we have  $\gamma_1^* = 0$ , whereas under asymmetric quadratic loss, we have  $\gamma_1^* = .25$ . To gauge the power of the tests, we need to construct a forecast that is suboptimal under *all* loss functions, not just under MSE or some other given loss function. One such set of suboptimal forecasts is obtained by simply adding independent noise to the optimal forecast,

$$\begin{aligned}
 \hat{Y}_{t+1,t} &= \hat{Y}_{t+1,t}^* + \xi \varepsilon_{t+1}, \\
 \varepsilon_{t+1} &\sim \text{iidN}(0, 1).
 \end{aligned}
 \tag{22}$$

This specification is representative of a number of sources of suboptimality. For example, the noise could come from including irrelevant variables in a prediction model or from random “judgmental adjustments” of a statistical forecasting model. We consider five values for the standard deviation of the noise:  $\xi = 0, .25, .5, .75$ , and 1. Of course, the case where  $\xi = 0$  corresponds to the forecast being optimal under MSE or asymmetric quadratic loss and thus can be used to examine the finite-sample size of the tests.

We study five tests of forecast optimality. The first test is the standard Mincer–Zarnowitz (1969) test, based on the regression

$$Y_{t+1} = \beta_0 + \beta_1 \hat{Y}_{t+1,t} + u_{t+1},$$

with the null hypothesis being  $\beta_0 = 0 \cap \beta_1 = 1$ . This test is appropriate for testing optimality under MSE loss but may over-reject forecasts that are optimal under other loss functions. For

this test, and the other two regression-based tests, we used the robust standard errors of Newey and West (1987).

The second test is another test of optimality under MSE loss but based on the forecast errors,  $e_{t+1,t} \equiv Y_{t+1} - \hat{Y}_{t+1,t}$ ,

$$e_{t+1,t} = \beta_0 + \beta_1 \hat{Y}_{t+1,t} + \beta_2 e_{t,t-1} + u_{t+1}.$$

In this case the null hypothesis is that  $\beta_0 = \beta_1 = \beta_2 = 0$ . The third test is our new test based on the indicator variable  $I_{t+1,t} \equiv \mathbf{1}(Y_{t+1} \leq \hat{Y}_{t+1,t})$ . This is used to test optimality under some unknown homogeneous loss function and is based on the regression

$$I_{t+1,t} = \beta_0 + \beta_1 \hat{Y}_{t+1,t} + \beta_2 I_{t,t-1} + u_{t+1}.$$

As shown in Proposition 2, the indicator variable  $I_{t+1,t}$  should be independent of all  $Z_t \in \mathcal{F}_t$ , although it has unknown mean. Thus the null hypothesis for this test is that  $\beta_1 = \beta_2 = 0$ .

The final two tests are based on spline-based estimation of the unknown loss function. The first test assumes that  $L(y, \hat{y}) = L(e)$  and approximates  $L$  with a quadratic spline using three nodes, so there are three free parameters. The nodes used in the simulation are  $\{\widehat{\text{Median}}[e_t|e_t < 0], 0, \widehat{\text{Median}}[e_t|e_t > 0]\}$ , where  $\widehat{\text{Median}}$  is the sample median. The instruments used in the moment conditions are a constant, the current value of the forecast, and the lagged values of the forecast error, the predicted variable, and the derivative of the loss function,  $\lambda$  (yielding a total of five moment conditions).

The second spline-based test allows  $L$  to depend on both  $y$  and  $\hat{y}$  and approximates  $L$  with a quadratic spline using three nodes and the specification given in (16), implying six free parameters. The nodes used are the same as for the simple spline-based test. The instruments used in the moment conditions are a constant, the current value of the forecast, and two lags each of the forecast error, the predicted variable, and the derivative of the loss function,  $\lambda$  (yielding a total of eight moments). All tests are conducted at the .05 level. The results are presented in Tables 2 and 3. We generated 3,000 replications of each simulation design.

Under MSE loss, all five tests have generally satisfactory size properties. The rejection frequencies are slightly high for  $T = 100$ , although this is commonly observed in tests based on robust standard errors. The rejection frequencies improve as the sample size increases, although the size of the test based on the flexible spline model (“spline 2”) remains slightly high even for  $T = 1,000$ .

Turning to the relative power of the tests, we find that under MSE loss the indicator-based test has approximately the same power as the Mincer–Zarnowitz (MZ) test. This occurs even though the indicator-based test is designed to detect suboptimality under any homogeneous loss function and may reflect, as noted in Section 1, that the MZ regression tests only a weak implication of forecast optimality. This interpretation is consistent with the finding that the test based on forecast errors (“ $e$  test”) has better power than both the MZ test and the indicator-based test, because it uses additional instruments, such as the lagged forecast error, that can identify serial correlation in the forecast errors. This effect is also likely to explain the good power of the two spline tests, which use even more conditioning variables as instruments.

Table 2. Finite-sample size and power of the tests under MSE loss

T	$\xi$	Test type				
		MZ test	$e$ test	$I$ test	Spline 1	Spline 2
100	0	.11	.12	.09	.09	.11
	.25	.23	.39	.24	.70	.70
	.5	.76	.92	.76	.91	.91
	.75	.98	1.00	.97	.98	.98
	1	1.00	1.00	1.00	.99	.99
250	0	.07	.09	.07	.06	.10
	.25	.28	.56	.34	.89	.88
	.5	.95	.99	.97	1.00	1.00
	.75	1.00	1.00	1.00	1.00	1.00
	1	1.00	1.00	1.00	1.00	1.00
1,000	0	.06	.06	.06	.06	.09
	.25	.66	.97	.80	1.00	1.00
	.5	1.00	1.00	1.00	1.00	1.00
	.75	1.00	1.00	1.00	1.00	1.00
	1	1.00	1.00	1.00	1.00	1.00

NOTE: This table reports the results of a Monte Carlo study of the finite-sample properties of the tests considered in this article. The DGP is a conditionally Gaussian AR(1)–GARCH(1, 1) process, and the sample sizes considered are denoted by  $T$ . When  $\xi = 0$ , the forecast is truly optimal under MSE loss, and as  $\xi$  grows, the forecast becomes increasingly contaminated with noise. The number of replications was 3,000.

Under asymmetric loss, the results are quite different. Unsurprisingly, the MZ and  $e$  tests reject the null hypothesis of optimality far more frequently than the nominal size. Even with just 100 observations, the MZ and  $e$  test reject the null hypothesis more than 60% of the time. The indicator-based test and the “spline 1” test both have good size and power properties. The more flexible spline-based test has poor size properties for  $T = 100$  and 250 (the empirical rejection frequencies are .45 and .35, compared with the nominal size of .05), and even for  $T = 1,000$ , the size is still large (.17).

Table 3. Finite-sample size and power of the tests under asymmetric loss

T	$\xi$	Test type				
		MZ test	$e$ test	$I$ test	Spline 1	Spline 2
100	0	.64	.62	.11	.12	.45
	.25	.65	.73	.23	.54	.69
	.5	.89	.96	.75	.95	.91
	.75	.98	1.00	.97	1.00	.98
	1	1.00	1.00	1.00	1.00	.99
250	0	.95	.93	.07	.08	.35
	.25	.96	.96	.31	.87	.88
	.5	.99	1.00	.96	1.00	.99
	.75	1.00	1.00	1.00	1.00	1.00
	1	1.00	1.00	1.00	1.00	1.00
1,000	0	1.00	1.00	.06	.05	.17
	.25	1.00	1.00	.79	1.00	1.00
	.5	1.00	1.00	1.00	1.00	1.00
	.75	1.00	1.00	1.00	1.00	1.00
	1	1.00	1.00	1.00	1.00	1.00

NOTE: This table reports the results of a Monte Carlo study of the finite-sample properties of the tests considered in this article. The DGP is a conditionally Gaussian AR(1)–GARCH(1, 1) process, and the sample sizes considered are denoted by  $T$ . When  $\xi = 0$ , the forecast is truly optimal under “quad–quad” loss, and as  $\xi$  grows, the forecast becomes increasingly contaminated with noise. The number of replications was 3,000.

Overall, these simulation results indicate that if the true loss function is MSE, then the MZ test and  $e$  test have reasonable size and power (with the latter test more powerful than the former) in finite samples. Both tests are slightly oversized when  $T \leq 250$ . The indicator-based test and the simple spline test have reasonable properties in finite samples for both MSE and non-MSE loss. The flexible spline test appears to require large samples ( $T \geq 1,000$ ) before the test's size is close to its nominal value, and thus rejections obtained using this test must be interpreted with caution.

With regard to our empirical application, the simulation results lend support to our interpretation of the reported findings. The reasonable finite-sample size of the regression and indicator-based tests support our rejection of the joint hypothesis of forecast optimality under conditional mean–variance dynamics (Assumption D2) and loss that is homogeneous in the forecast error (Assumption L2). Similarly, our rejection of the simple error-based spline model and our failure to reject the more general spline model that allows loss to depend on both the forecast error and the level of the outcome variable is not overturned by the tendency of tests based on the latter to be oversized.

## 5. CONCLUSION

Motivated by the surprising rejection of the optimality of the Fed's internal forecasts (i.e., the “Greenbook” forecasts) of real GDP growth under MSE loss, in this article we have proposed new tests of forecast optimality that are applicable when the forecaster's loss function is unknown. Our first set of tests applies when the loss function is homogeneous in the forecast error and the variable of interest has dynamics in the conditional mean and variance but constant higher-order moments. This restriction on the loss function is quite weak, whereas the necessary restrictions on the DGP can be easily tested. In particular, we propose a new set of quantile tests that does not require knowledge of the conditional mean and variance of the predicted variable, both of which depend on the forecaster's (unobserved) information set.

Our second set of tests is based on flexible parametric approximations to the unknown loss function. We found significant evidence against the optimality of the Fed's forecasts when the loss function was assumed to be a function solely of the forecast error, but found no evidence against optimality when the loss function was allowed to depend on the forecast and the realization separately. Our estimates of the loss function suggest that the Fed issues “conservative” estimates of economic growth, with underestimates penalized less heavily than overestimates. This conservatism appears to be particularly important when economic growth is moderate or low. This is consistent with the Fed viewing overpredictions of economic growth as being not just more costly than underpredictions, but also disproportionately more costly in times of low growth, perhaps because such forecasts may incorrectly signal a recovery from a recession and could result in an overly tight monetary policy at a critical point in time.

## APPENDIX: PROOFS

### Proof of Proposition 1

The proof of part a was shown by Granger (1969) and Christoffersen and Diebold (1997). It is similar to our proof of Proposition 2 herein.

For part b, given the representation result in part a, we know that the forecast error associated with the optimal forecast is  $e_{t+h,t}^* = Y_{t+h} - \hat{Y}_{t+h,t}^* = \varepsilon_{t+h} - \alpha_h^*$ , where  $\alpha_h^*$  solves  $\min_{\alpha} \int L(\varepsilon_{t+h} - \alpha) dF_{\varepsilon,h}$ . Because  $\alpha_h^*$  is constant for fixed  $h$ , we have that  $e_{t+h,t}^*$  is independent of all  $Z_t \in \mathcal{F}_t$ .

For part c, consider  $h > 0$  and  $j > 0$ . Let

$$Y_{t+h+j} = E_t[Y_{t+h+j}] + \eta_{t+h+j}, \eta_{t+h+j} | \mathcal{F}_t \sim F_{\varepsilon,h+j}(0, \sigma_{\varepsilon,h+j}^2)$$

and

$$Y_{t+h+j} = E_{t+j}[Y_{t+h+j}] + \varepsilon_{t+h+j}, \varepsilon_{t+h+j} | \mathcal{F}_t \sim F_{\varepsilon,h}(0, \sigma_{\varepsilon,h}^2).$$

Also let  $\sigma_{\varepsilon,h}^2 < \infty$ , and further assume that  $\sigma_{\varepsilon,h+j}^2 < \infty$ . Using again the fact that  $\hat{Y}_{t+h,t}^* = E_t[Y_{t+h}] + \alpha_h^*$ , we have that  $e_{t+h+j,t}^* = \eta_{t+h+j} - \alpha_{h+j}^*$  and  $e_{t+h+j,t+j}^* = \varepsilon_{t+h+j} - \alpha_{h+j}^*$ , where  $\alpha_h^*$  and  $\alpha_{h+j}^*$  are constants. Thus  $V_t[e_{t+h+j,t}^*] = V_t[\eta_{t+h+j}] = \sigma_{\varepsilon,h+j}^2$  and  $V_t[e_{t+h+j,t+j}^*] = \sigma_{\varepsilon,h}^2$ , where these moments are independent of  $t$  by Assumption D1. Also note that  $V[e_{t+h+j,t}^*] = E[E_t[\eta_{t+h+j}^2]] = \sigma_{\varepsilon,h+j}^2$  and, similarly,  $V[e_{t+h+j,t+j}^*] = \sigma_{\varepsilon,h}^2$ . Now we seek to show that  $\sigma_{\varepsilon,h+j}^2 \geq \sigma_{\varepsilon,h}^2$ ,

$$\begin{aligned} V[e_{t+h+j,t}^*] &= V_t[Y_{t+h+j} - E_t[Y_{t+h+j}]] \\ &= V_t[\varepsilon_{t+h+j} + (E_{t+j}[Y_{t+h+j}] - E_t[Y_{t+h+j}])] \\ &= \sigma_{\varepsilon,h+j}^2 + V_t[E_{t+j}[Y_{t+h+j}]] \\ &\quad + 2 \operatorname{cov}_t[\varepsilon_{t+h+j}, E_{t+j}[Y_{t+h+j}] - E_t[Y_{t+h+j}]] \\ &\geq \sigma_{\varepsilon,h+j}^2 \\ &= V[e_{t+h,t}^*]. \end{aligned}$$

The first equality follows from the equality of the conditional and unconditional variance of the forecast error under Assumption D1, the third equality follows from the fact that  $E_t[Y_{t+h+j}]$  is constant given  $\mathcal{F}_t$ , the weak inequality follows from the nonnegativity of  $V_t[E_{t+j}[Y_{t+h+j}]]$  and  $E_{t+j}[\varepsilon_{t+h+j} \cdot \phi(Z_{t+j})] = 0$ , and the final equality follows from the fact that  $F_{\varepsilon,h}$  does not change with  $t$ . The cases where  $h = 0$  and/or  $j = 0$  are trivial. Thus  $V[e_{t+h+j,t}^*] \geq V[e_{t+h,t}^*] \forall h, j \geq 0$  if  $V[e_{t+h+j,t}^*] < \infty$ . If  $\sigma_{\varepsilon,h}^2 < \infty$  but  $\sigma_{\varepsilon,h+j}^2$  is infinite, then the proposition holds trivially.

### Proof of Proposition 2

To prove part a, by homogeneity, we have

$$\begin{aligned} \hat{Y}_{t+h,t}^* &\equiv \arg \min_{\hat{y}} \int L(y - \hat{y}) dF_{t+h,t}(y) \\ &= \arg \min_{\hat{y}} \int \left[ g\left(\frac{1}{\sigma_{t+h,t}}\right) \right]^{-1} L\left(\frac{1}{\sigma_{t+h,t}}(y - \hat{y})\right) dF_{t+h,t}(y) \\ &= \arg \min_{\hat{y}} \int L\left(\frac{1}{\sigma_{t+h,t}}(y - \hat{y})\right) dF_{t+h,t}(y) \\ &= \arg \min_{\hat{y}} \int L\left(\frac{1}{\sigma_{t+h,t}}(\mu_{t+h,t} + \sigma_{t+h,t}\eta_{t+h} - \hat{y})\right) dF_{\eta,h}(\eta). \end{aligned}$$

We represent a forecast as  $\hat{Y}_{t+h,t} = \mu_{t+h,t} + \sigma_{t+h,t} \cdot \hat{\gamma}_{t+h,t}$ , so that

$$\begin{aligned} \hat{Y}_{t+h,t}^* &= \mu_{t+h,t} \\ &\quad + \sigma_{t+h,t} \cdot \arg \min_{\hat{\gamma}} \int L\left(\frac{1}{\sigma_{t+h,t}}(\mu_{t+h,t} + \sigma_{t+h,t}\eta_{t+h} \right. \\ &\quad \left. - \mu_{t+h,t} - \sigma_{t+h,t}\hat{\gamma})\right) dF_{\eta,h}(\eta) \end{aligned}$$

$$\begin{aligned}
&= \mu_{t+h,t} + \sigma_{t+h,t} \cdot \arg \min_{\hat{\gamma}} \int L(\eta_{t+h} - \hat{\gamma}) dF_{\eta,h}(\eta) \\
&= \mu_{t+h,t} + \sigma_{t+h,t} \cdot \gamma_h^*,
\end{aligned}$$

where the last line follows from the fact that  $F_{\eta,h}$  is time-invariant under Assumption D2.

The proof of part b follows from noting that  $\hat{d}_{t+h,t}^* = \eta_{t+h} - \gamma_h^*$ , where  $\gamma_h^*$  is a constant for fixed  $h$ , and, by Assumption D2,  $\eta_{t+h}$  is independent of all elements in  $\mathcal{F}_t$  and has unit variance.

### Proof of Corollary 1

Under Assumptions D2' and L2, we have, from Proposition 2, that  $\hat{Y}_{t+h,t}^* = \sigma_{t+h,t}(\beta + \gamma_h^*)$ . Thus  $\hat{d}_{t+h,t}^* \equiv e_{t+h,t}^*/\hat{Y}_{t+h,t}^* = (\eta_{t+h} - \gamma_h^*)/(\beta + \gamma_h^*)$ , that is, an affine transformation of  $\eta_{t+h}$ . The result follows by noting that  $\eta_{t+h}$  is independent of all  $\mathbf{Z}_t \in \mathcal{F}_t$ .

### Proof of Proposition 3

To prove part a under Assumptions D1 and L1 or D2 and L2, we know from Propositions 1 and 2 that  $Y_{t+h,t}^* = \mu_{t+h,t} + \sigma_{t+h,t} \cdot \gamma_h^*$ , with  $\sigma_{t+h,t}$  constant under Assumption D1, where  $\gamma_h^*$  depends only on the loss function and  $F_{\eta,h}$ . Now note that  $F_{t+h,t}(\hat{Y}_{t+h,t}^*) \equiv \Pr[Y_{t+h} \leq \hat{Y}_{t+h,t}^* | \mathcal{F}_t] = \Pr[\mu_{t+h,t} + \sigma_{t+h,t}\eta_{t+h} \leq \mu_{t+h,t} + \sigma_{t+h,t} \cdot \gamma_h^* | \mathcal{F}_t] = \Pr[\eta_{t+h} \leq \gamma_h^* | \mathcal{F}_t]$ , which is constant under Assumption D2. If we denote  $q_h^* \equiv \Pr[\eta_{t+h} \leq \gamma_h^* | \mathcal{F}_t]$ , then  $\hat{Y}_{t+h,t}^*$  is the  $q_h^*$  conditional quantile of  $Y_{t+h} | \mathcal{F}_t \forall t$ . Note that  $q_h^*$  is a function only of the loss function and  $F_{\eta,h}$ .

To prove part b, because  $I_{t+h,t}^*$  is a binary random variable and  $\Pr[I_{t+h,t}^* = 1 | \mathcal{F}_t] = \Pr[Y_{t+h} \leq \hat{Y}_{t+h,t}^* | \mathcal{F}_t] = q_h^* \forall t$ , we thus have that  $I_{t+h,t}^*$  is independent of all  $\mathbf{Z}_t \in \mathcal{F}_t$ .

### Proof of Proposition 4

To prove part a, following the steps in the proof of Proposition 2, we find that

$$\begin{aligned}
\hat{Y}_{t+h,t}^* &= \mu_{t+h,t} + \arg \min_{\hat{\alpha}} \int L(\sigma_{t+h,t}(\eta_{t+h} - \hat{\alpha})) dF_{\eta,h}(\eta) \\
&\equiv \mu_{t+h,t} + \alpha_{t+h,t}^*,
\end{aligned}$$

where  $\alpha_{t+h,t}^*$  depends on the loss function,  $L$ ; the conditional standard deviation of  $Y_{t+h}$ ,  $\sigma_{t+h,t}$ ; and the distribution of the innovation,  $F_{\eta,h}$ .

For part b, similarly,

$$\begin{aligned}
\hat{Y}_{t+h,t}^* &= \mu_{t+h,t} + \sigma_{t+h,t} \cdot \arg \min_{\hat{\gamma}} \int L(\eta_{t+h} - \hat{\gamma}) dF_{\eta,t+h,t}(\eta) \\
&\equiv \mu_{t+h,t} + \sigma_{t+h,t} \cdot \gamma_{t+h,t}^*.
\end{aligned}$$

Thus  $\gamma_{t+h,t}^*$  will be a function of the loss function and  $F_{\eta,t+h,t}$ , with the latter depending on time-varying properties of the conditional distribution of  $Y_{t+h} | \mathcal{F}_t$  beyond the conditional mean and variance.

[Received April 2006. Revised August 2006.]

## REFERENCES

- Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2001), "The Distribution of Realized Exchange Rate Volatility," *Journal of the American Statistical Association*, 96, 42–55.
- (2003), "Modeling and Forecasting Realized Volatility," *Econometrica*, 71, 579–625.
- Bollerslev, T. (1986), "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 31, 307–327.
- Campbell, S. D., and Diebold, F. X. (2005), "Weather Forecasting for Weather Derivatives," *Journal of the American Statistical Association*, 100, 6–16.
- Capistran, C. (2006), "Bias in Federal Reserve Inflation Forecasts: Is the Federal Reserve Irrational or Just Cautious?" manuscript, Banco de Mexico.
- Christoffersen, P. F. (1998), "Evaluating Interval Forecasts," *International Economic Review*, 39, 841–862.

- Christoffersen, P. F., and Diebold, F. X. (1997), "Optimal Prediction Under Asymmetric Loss," *Econometric Theory*, 13, 808–817.
- Clements, M. P., and Hendry, D. F. (1998), *Forecasting Economic Time Series*, Cambridge, U.K.: Cambridge University Press.
- (2006), "Forecasting With Breaks in Data Processes," in *Handbook of Economic Forecasting*, eds. C. W. J. Granger, G. Elliott, and A. Timmermann, Amsterdam: North-Holland, pp. 605–657.
- Croushore, D., and Stark, T. (2001), "A Real-Time Data Set for Macroeconomists," *Journal of Econometrics*, 105, 111–130.
- Diebold, F. X., and Lopez, J. (1996), "Forecast Evaluation and Combination," in *Handbook of Statistics*, eds. G. S. Maddala and C. R. Rao, Amsterdam: North-Holland, pp. 241–268.
- Diebold, F. X., and Rudebusch, G. D. (1991), "Forecasting Output With the Composite Leading Index: A Real-Time Analysis," *Journal of the American Statistical Association*, 86, 603–610.
- Elliott, G., Komunjer, I., and Timmermann, A. (2005), "Estimation and Testing of Forecast Rationality Under Flexible Loss," *Review of Economic Studies*, 72, 1107–1125.
- Engle, R. F. (1982), "Autoregressive Conditional Heteroskedasticity With Estimates of the Variance of U.K. Inflation," *Econometrica*, 50, 987–1008.
- Engle, R. F., Lilien, D. M., and Robins, R. P. (1987), "Estimating Time-Varying Risk Premia in the Term Structure: The ARCH-M Model," *Econometrica*, 55, 391–407.
- Garratt, A., Lee, K., Pesaran, M. H., and Shin, Y. (2003), "Forecast Uncertainties in Macroeconomic Modeling: An Application to the U.K. Economy," *Journal of the American Statistical Association*, 464, 829–838.
- Granger, C. W. J. (1969), "Prediction With a Generalized Cost Function," *OR*, 20, 199–207.
- Granger, C. W. J., and Newbold, P. (1986), *Forecasting Economic Time Series* (2nd ed.), New York: Academic Press.
- Hall, A. R. (2005), *Generalized Method of Moments*, Oxford, U.K.: Oxford University Press.
- Hong, Y. (1999), "Hypothesis Testing in Time Series via the Empirical Characteristic Function: A Generalized Spectral Density Approach," *Journal of the American Statistical Association*, 94, 1201–1220.
- Jarque, C. M., and Bera, A. K. (1987), "A Test for Normality of Observations and Regression Residuals," *International Statistical Review*, 55, 163–172.
- Koenker, R., and Bassett, G. (1978), "Regression Quantiles," *Econometrica*, 46, 33–50.
- McConnell, M. M., and Perez-Quiros, G. (2000), "Output Fluctuations in the United States: What Has Changed Since the Early 1980s?" *American Economic Review*, 90, 1464–1476.
- Mincer, J., and Zarnowitz, V. (1969), "The Evaluation of Economic Forecasts," in *Economic Forecasts and Expectations*, ed. J. Mincer, New York: National Bureau of Economic Research, pp. 81–111.
- Newey, W. K., and McFadden, D. (1994), "Large-Sample Estimation and Hypothesis Testing," in *Handbook of Econometrics*, Vol. IV, eds. R. F. Engle and D. McFadden, Amsterdam: North-Holland, pp. 2111–2245.
- Newey, W. K., and West, K. D. (1987), "A Simple Positive Semi-Definite, Heteroskedastic and Autocorrelation-Consistent Covariance Matrix," *Econometrica*, 55, 703–708.
- Nobay, A. R., and Peel, D. A. (2003), "Optimal Discretionary Monetary Policy in a Model of Asymmetric Central Bank Preferences," *Economic Journal*, 113, 657–665.
- Patton, A. J., and Timmermann, A. (2006), "Properties of Optimal Forecasts Under Asymmetric Loss and Nonlinearity," *Journal of Econometrics*, in press.
- Romer, C. D., and Romer, D. H. (2000), "Federal Reserve Information and the Behavior of Interest Rates," *American Economic Review*, 90, 429–457.
- Stock, J. H., and Watson, M. W. (2002), "Has the Business Cycle Changed and Why?" in *NBER Macroeconomics Annual 2002*, eds. M. Gertler and K. Rogoff, Cambridge, MA: MIT Press, pp. 159–218.
- Taylor, S. J. (1982), "Financial Returns Modelled by the Product of Two Stochastic Processes: A Study of Daily Sugar Prices, 1961–79," in *Time Series Analysis: Theory and Practice 1*, ed. O. D. Anderson, Amsterdam: North-Holland, pp. 203–226.
- Varian, H. R. (1974), "A Bayesian Approach to Real Estate Assessment," in *Studies in Bayesian Econometrics and Statistics in Honor of Leonard J. Savage*, eds. S. E. Fienberg and A. Zellner, Amsterdam: North-Holland, pp. 195–208.
- Weiss, A. A. (1996), "Estimating Time Series Models Using the Relevant Cost Function," *Journal of Applied Econometrics*, 11, 539–560.
- West, K. D., Edison, H. J., and Cho, D. (1993), "A Utility-Based Comparison of Some Models of Exchange Rate Volatility," *Journal of International Economics*, 35, 23–46.
- Zellner, A. (1986), "Bayesian Estimation and Prediction Using Asymmetric Loss Functions," *Journal of the American Statistical Association*, 81, 446–451.