

New Tests of Forecast Optimality Across Multiple Horizons*

Andrew J. Patton

Allan Timmermann

Duke University

University of California, San Diego

March 26, 2010

PRELIMINARY AND INCOMPLETE.

Abstract

We propose new joint tests of forecast optimality that exploit information contained in multi-horizon forecasts. In addition to implying zero bias and zero autocorrelation in forecast errors, we show that forecast optimality under squared error loss also implies testable restrictions on second moments of the data ordered across forecast horizons. In particular, the variance of the forecast error should be increasing in the horizon; the variance of the forecast itself should be decreasing in the horizon; and the variance of forecast revisions should be bounded by twice the covariance of revisions with the target variable. These bounds on second moments can be restated as inequality constraints in a regression framework and tested using the approach of Wolak (1989). Moreover, some of the proposed tests can be conducted without the need for data on the target variable, which is particularly useful in the presence of large measurement errors. We also propose a new univariate optimal revision test that constrains the coefficients in a regression of the target variable on the long-horizon forecast and the sequence of interim forecast revisions. The size and power of the new tests are compared with those of extant tests through Monte Carlo simulations. An empirical application to the Federal Reserve's Greenbook forecasts is used to illustrate the tests.

Keywords: Forecast optimality, real-time data, variance bounds, survey forecasts, forecast horizon.

J.E.L. Codes: C53, C22, C52.

*We thank Tim Bollerslev and Ken West as well as seminar participants at Duke, UCSD, EC^2 conference on real-time econometrics (December, 2009), and the 6th forecasting symposium at the ECB.

1 Introduction

Forecasts recorded at multiple horizons, for example from one to several quarters into the future, are becoming increasingly common in empirical practice. For example, the surveys conducted by the Philadelphia Federal Reserve (Survey of Professional Forecasters), Consensus Economics or Blue Chip and the forecasts produced by the IMF (World Economic Outlook), the Congressional Budget office, the Bank of England and the Board of the Federal Reserve all cover several horizons. Similarly, econometric models are commonly used to generate multi-horizon forecasts, see, e.g., Faust and Wright (2009), Marcellino, Stock and Watson (2006), and Clements (1997). With the availability of such multi-horizon forecasts, there is a growing need for tests of optimality to exploit the information in the complete “term structure” of forecasts recorded across all horizons. By simultaneously exploiting information across several horizons, rather than focusing separately on individual horizons, multi-horizon forecast tests offer the potential of drawing more powerful conclusions about the ability of forecasters to produce optimal forecasts. This paper derives a number of novel and simple implications of forecast optimality and compares tests based on these implications with extant tests.

A well-known implication of forecast optimality is that, under squared error loss, the mean squared forecast error should be a non-decreasing function of the forecast horizon, see, e.g., Diebold (2001) and Patton and Timmermann (2007a). A similar property holds for the forecasts themselves: Internal consistency of a sequence of optimal forecasts implies that the variance of the forecasts should be a non-increasing function of the forecast horizon. Intuitively, this property holds because, just as the variance of the realized value must be (weakly) greater than the variance of its conditional expectation, the variance of the expectation conditional on a large information set (corresponding to a short horizon) must exceed that of the expectation conditional on a smaller information set (corresponding to a long horizon). It is also possible to show that optimal updating of forecasts implies that the variance of the forecast revision should exceed twice the covariance between the forecast revision and the actual value. It is uncommon to test such variance bounds in empirical practice, in part due to the difficulty in setting up joint tests of these bounds. We suggest and illustrate testing these monotonicity properties via tests of inequality constraints using the methods of Gouriéroux et al. (1982) and Wolak (1987, 1989).

Tests of forecast optimality have conventionally been based on comparing predicted and “re-

alized” values of the outcome variable. This severely constrains inference in some cases since, as shown by Croushore (2006), Croushore and Stark (2001) and Corradi, Fernandez and Swanson (2009), revisions to macroeconomic variables can be very considerable. This raises questions that can be difficult to address such as “what are the forecasters trying to predict?”, i.e. first-release data or final revisions. We show that variations on both the new and extant optimality tests can be applied without the need for observations on the target variable. These tests are particularly useful in situations where the target variable is not observed (such as for certain types of volatility forecasts) or is measured with considerable noise (as in the case of output forecasts).

Conventional tests of forecast optimality regress the realized value of the predicted variable on an intercept and the forecast for a single horizon and test the joint implication that the intercept and slope coefficient are zero and one, respectively (Mincer and Zarnowitz (1969).) In the presence of forecasts covering multiple horizons, we show that a complete test that imposes internal consistency restrictions on the forecast revisions gives rise to a generalized efficiency regression. Using a single equation, this test is undertaken by regressing the realized value on an intercept, the long-horizon forecast and the sequence of intermediate forecast revisions. A set of zero-one equality restrictions on the intercept and slope coefficients are then tested. A key difference from the conventional Mincer-Zarnowitz test is that the joint consistency of all forecasts at different horizons is tested by this generalized regression.

Analysis of forecast optimality is usually predicated on covariance stationarity assumptions. However, we show that the conventional assumption that the target variable and forecast are (jointly) covariance stationary is not needed for some of our tests and can be relaxed provided that forecasts for different horizons are lined up in “event time”, as studied by Nordhaus (1987) and Clements (1997). In particular, we show that the second moment bounds continue to hold in the presence of structural breaks in the variance of the innovation to the predicted variable. We present a general family of data generating processes for which the variance bounds continue to hold.

To shed light on the statistical properties of the variance bound and regression-based tests of forecast optimality, we undertake a set of Monte Carlo simulations. These simulations consider various scenarios with zero, low and high measurement error in the predicted variable and deviations from forecast optimality in a variety different directions. We find that the covariance bound and the single-equation test of joint forecast consistency have good power and size properties. Specifically, they are generally better than conventional Mincer-Zarnowitz tests conducted for individual hori-

zons which either tend to be conservative—if a Bonferroni bound is used to summarize the evidence across multiple horizons—or suffer from substantial size distortions, if the multi-horizon regressions are estimated as a system. Our simulations suggest that the various bounds and regression tests have complementary properties in the sense that they have power in different directions and so can identify different types of suboptimal behavior among forecasters.

An empirical application to Greenbook forecasts of GDP growth, changes to the GDP deflator and consumer price inflation confirms the findings from the simulations. In particular, we find that conventional regression tests often fail to reject the null of forecast optimality. In contrast, the new variance-bounds tests and single equation multi-horizon tests have better power and are able to identify deviations from forecast optimality.

The outline of the paper is as follows. Section 2 presents some novel implications of optimality of forecasts across multiple horizons, and describes hypothesis tests associated with these implications. Section 3 considers regression-based tests of forecast optimality and Section 4 discusses the role of stationarity for fixed-event forecasts. Section 5 presents the results from the Monte Carlo study, while Section 6 provides an empirical application to Federal Reserve Greenbook forecasts. Section 7 concludes.

2 Variance Bounds Tests

In this section we derive variance and covariance bounds that can be used to test the optimality of a sequence of forecasts recorded at different horizons. These are presented as corollaries to the well-known theorem that the optimal forecast under quadratic loss is the conditional mean. The proofs of these corollaries are straightforward, and are collected in the Appendix.

2.1 Assumptions and background

Consider a univariate time series, $Y \equiv \{Y_t; t = 1, 2, \dots\}$, and suppose that forecasts of this variable are recorded at different points in time, $t = 1, \dots, T$ and at different horizons, $h = h_1, \dots, h_H$. Forecasts of Y_t made h periods previously will be denoted as $\hat{Y}_{t|t-h}$ and is thus conditioned on the available information set at time $t - h$, \mathcal{F}_{t-h} , which is taken to be the σ -field generated by $\{\tilde{Z}_{t-h-k}; k \geq 0\}$, where \tilde{Z}_{t-h} is a vector of predictor variables capturing elements in the forecaster's information set at time $t - h$. Note that the target variable, Y_t , may or may not be an element of

\tilde{Z}_t , depending on whether this variable is observable to the forecaster or not. Forecast errors are given by $e_{t|t-h} = Y_t - \hat{Y}_{t|t-h}$. We consider an $(H \times 1)$ vector of multi-horizon forecasts for horizons $h_1 < h_2 < \dots < h_H$, with generic long and short horizons denoted by h_L and h_S ($h_L > h_S$). Note that the forecast horizons, h_i , can be positive, zero or negative, corresponding to forecasting, nowcasting or backcasting, and further note that we do not require the forecast horizons to be equally spaced.

We consider tests of forecast optimality under the assumption that the forecaster has squared error loss, and under that assumption we have the following well-known theorem, see Granger (1969) for example.

Theorem 1 (Optimal forecast under MSE) *Assume that the forecaster's loss function is $L(y, \hat{y}) = (y - \hat{y})^2$ and that the conditional mean of the target variable, $E[Y_t | \mathcal{F}_{t-h}]$ is a.s. finite for all t . Then*

$$\hat{Y}_{t|t-h}^* \equiv \arg \min_{\hat{y} \in \mathcal{Y}} E[(Y_t - \hat{y})^2 | \mathcal{F}_{t-h}] = E[Y_t | \mathcal{F}_{t-h}] \quad (1)$$

where $\mathcal{Y} \subseteq \mathbb{R}$ is the set of possible values for the forecast.

From this result it is simple to show that the associated forecast errors, $e_{t|t-h}^* = Y_t - \hat{Y}_{t|t-h}^*$ are mean-zero and uncorrelated with any $Z_{t-h} \in \mathcal{F}_{t-h}$. We next describe a variety of forecast optimality tests based on corollaries to this theorem.

Our analysis does not restrict the predicted “event” to be a single period outcome such as GDP growth in 2011Q4. Instead the predicted outcome could be the cumulated GDP growth over some period, say 2001Q1 through 2011Q4. Only the interpretation of the forecast horizon will change in the latter situation, e.g. if the point of the forecast is 2011Q3, in which case part of the predicted variable may be observed.

The tests proposed and studied in this paper take the forecasts as primitive, and if the forecasts are generated by particular econometric models, rather than by a combination of modeling and judgemental information, the estimation error embedded in those models is ignored. In the presence of estimation error the results established here need not hold. In practice forecasters face parameter uncertainty, model uncertainty and model instability issues and, as shown by West and McCracken (1998), parameter estimation error can lead to substantial skews in unadjusted t -statistics. While some of these effects can be addressed when comparing the relative precision of two forecasting models evaluated at the pseudo-true probability limit of the model estimates (West (1996)) or when

comparing forecasting methods conditionally (Giacomini and White (2006)), it is not in general possible to establish results for the absolute forecasting performance of a forecasting model. For example, under recursive parameter estimation, forecast errors will generally be serially correlated (Timmermann (1993)) and the mean squared forecast error may not be increasing in the forecast horizon (Schmidt (1974), Clements and Hendry (1998)). Existing analytical results are very limited, however, as they assume a particular model (e.g., an AR(1) specification), whereas in practice forecasts from surveys and forecasts reported by central banks reflect considerable judgmental information. We leave the important extension to incorporate estimation error to future research.

Some of the results derived below will make use of a standard covariance stationarity assumption:

Assumption S1: *The target variable, Y_t , is generated by a covariance stationary process.*

2.2 Monotonicity of mean squared errors

From forecast optimality under squared-error loss, (1), it follows that, for any $\tilde{Y}_{t|t-h} \in \mathcal{F}_{t-h}$,

$$E_{t-h} \left[\left(Y_t - \hat{Y}_{t|t-h}^* \right)^2 \right] \leq E_{t-h} \left[\left(Y_t - \tilde{Y}_{t|t-h} \right)^2 \right].$$

In particular, the optimal forecast at time $t - h_S$ must be at least as good as the forecast associated with a longer horizon, h_L :

$$E_{t-h} \left[\left(Y_t - \hat{Y}_{t|t-h_S}^* \right)^2 \right] \leq E_{t-h} \left[\left(Y_t - \hat{Y}_{t|t-h_L}^* \right)^2 \right].$$

This leads us to the first corollary to Theorem 1 (all proofs are contained in the Appendix):

Corollary 1 *Under the assumptions of Theorem 1 and S1, it follows that*

$$MSE(h_S) \equiv E \left[\left(Y_t - \hat{Y}_{t|t-h_S}^* \right)^2 \right] \leq E \left[\left(Y_t - \hat{Y}_{t|t-h_L}^* \right)^2 \right] \equiv MSE(h_L) \quad \text{for any } h_S < h_L.$$

Given a set of forecasts available at horizons $h_1 \leq h_2 \leq \dots \leq h_H$, it follows that the mean squared error (MSE) associated with an optimal forecast, $e_{t|t-h}^* \equiv Y_t - \hat{Y}_{t|t-h}^*$, is a non-decreasing function of the forecast horizon:

$$E \left[e_{t|t-h_1}^{*2} \right] \leq E \left[e_{t|t-h_2}^{*2} \right] \leq \dots \leq E \left[e_{t|t-h_H}^{*2} \right] \rightarrow V[Y_t] \quad \text{as } h_H \rightarrow \infty, \quad (2)$$

The inequalities are strict if more forecast-relevant information becomes available as the forecast horizon shrinks to zero.¹ This property is well-known and is discussed by, e.g., Diebold (2001) and

¹For example, for a non-degenerate AR(1) process the MSEs will strictly increase with the forecast horizon, while for an MA(1) the inequality will be strict only for $h = 1$ vs. $h = 2$, while for longer horizons the MSEs will be equal.

Patton and Timmermann (2007a).

Example 1: To illustrate a violation of this property, consider the case of a “lazy” forecaster, who, in constructing a short-horizon forecast, $\tilde{Y}_{t|t-h_S}$, does not update his long-horizon forecast, $\tilde{Y}_{t|t-h_L}$, with relevant information, and hides this lack of updating by adding a small amount of zero-mean, independent noise to the long-horizon forecast. In that case:

$$\tilde{Y}_{t|t-h_S} = \tilde{Y}_{t|t-h_L} + u_{t-h_S}, \quad u_{t-h_S} \perp \tilde{Y}_{t|t-h_L}, u_{t-h_S} \perp Y_t. \quad (3)$$

We then have

$$V[e_{t|t-h_S}] = V[Y_t - \tilde{Y}_{t|t-h_L} - u_{t-h_S}] = V[e_{t|t-h_L} - u_{t-h_S}] = V[e_{t|t-h_L}] + V[u_{t-h_S}] > V[e_{t|t-h_L}].$$

Hence the short-horizon forecast generates a larger MSE than the long-horizon forecast, revealing the sub-optimality of the short-horizon forecast.

2.3 Testing monotonicity in squared forecast errors

The results derived so far suggest testing forecast optimality via a test of the weak monotonicity in the “term structure” of mean squared errors, (2), to use the terminology of Patton and Timmermann (2008). This feature of rational forecasts is relatively widely known, but has generally not been used to test forecast optimality. Capistran (2007) is the only paper we are aware of that exploits this property to develop a test. His test is based on Bonferroni bounds, which are quite conservative in this application. Here we advocate an alternative procedure for testing non-decreasing MSEs at longer forecast horizons that is based on the inequalities in (2).

We consider ranking the MSE-values for a set of forecast horizons $h = h_1, h_2, \dots, h_H$. Denoting the expected (population) value of the MSEs by $\boldsymbol{\mu}^e = [\mu_1^e, \dots, \mu_H^e]'$, with $\mu_j^e \equiv E[e_{t|t-h_j}^2]$, and defining the associated MSE differentials as

$$\Delta_j^e \equiv \mu_j - \mu_{j-1} = E[e_{t|t-h_j}^2] - E[e_{t|t-h_{j-1}}^2],$$

we can rewrite the inequalities in (2) as

$$\Delta_j^e \geq 0, \quad \text{for } j = h_2, \dots, h_H. \quad (4)$$

Following earlier work on multivariate inequality tests in regression models by Gourieroux, et

al. (1982), Wolak (1987, 1989) proposed testing (weak) monotonicity through the null hypothesis:

$$\begin{aligned} H_0 &: \Delta^e \geq \mathbf{0}, \\ \text{vs. } H_1 &: \Delta^e \in \mathbb{R}^{H-1}, \end{aligned} \tag{5}$$

where the $(H-1) \times 1$ vector of MSE-differentials is given by $\Delta^e \equiv [\Delta_2^e, \dots, \Delta_H^e]'$. In contrast, Δ^e is unconstrained under the alternative. Tests can be based on the sample analogs $\hat{\Delta}_j^e = \hat{\mu}_j - \hat{\mu}_{j-1}$ for $\hat{\mu}_j \equiv \frac{1}{T} \sum_{t=1}^T e_{t|t-h_j}^2$. Wolak (1987, 1989) derives a test statistic whose distribution under the null is a weighted sum of chi-squared variables, $\sum_{i=1}^{H-1} \omega(H-1, i) \chi^2(i)$, where $\omega(H-1, i)$ are the weights and $\chi^2(i)$ is a chi-squared variable with i degrees of freedom. Approximate critical values for this test can be calculated through Monte Carlo simulation. For further description of this test and other approaches to testing multivariate inequalities, see Patton and Timmermann (2009).

2.4 Monotonicity of mean squared forecasts

We now present a novel implication of forecast optimality that can be tested when data on the target variable is not available or not reliable. Recall that, under optimality, $E_{t-h} [e_{t|t-h}^*] = 0$ which implies that $Cov [\hat{Y}_{t|t-h}^*, e_{t|t-h}^*] = 0$. Thus we obtain the following corollary:

Corollary 2 *Under the assumptions of Theorem 1 and S1, we have $V[Y_t] = V[\hat{Y}_{t|t-h}^*] + E[e_{t|t-h}^{*2}]$. From Corollary 1 we have $E[e_{t|t-h_S}^{*2}] \leq E[e_{t|t-h_L}^{*2}]$ for any $h_S < h_L$, which then yields*

$$V[\hat{Y}_{t|t-h_S}^*] \geq V[\hat{Y}_{t|t-h_L}^*] \quad \text{for any } h_S < h_L.$$

Further, since $E[\hat{Y}_{t|t-h}^*] = E[Y_t]$, we also obtain an inequality on the mean-squared forecasts:

$$E[\hat{Y}_{t|t-h_S}^{*2}] \geq E[\hat{Y}_{t|t-h_L}^{*2}] \quad \text{for any } h_S < h_L. \tag{6}$$

This reveals that a weakly increasing pattern in MSE-values as the forecast horizon increases implies a weakly *decreasing* pattern in the variance of the forecasts themselves. This simple result provides the surprising implication that (one aspect of) forecast optimality may be tested *without* the need for a measure of the target variable. A test of this implication can again be based on Wolak's (1989) approach by defining the vector $\Delta^f \equiv [\Delta_2^f, \dots, \Delta_H^f]'$, where $\Delta_j^f \equiv E[\hat{Y}_{t|t-h_j}^{*2}] - E[\hat{Y}_{t|t-h_{j-1}}^{*2}]$ and testing the null hypothesis that differences in mean squared forecasts are weakly

negative for all forecast horizons:

$$\begin{aligned} H_0 &: \Delta^f \leq \mathbf{0}, \\ \text{vs. } H_1 &: \Delta^f \in \mathbb{R}^{H-1}. \end{aligned} \tag{7}$$

It is worth pointing out some limitations to this type of test. Tests that do not rely on observing the realized values of the target variable are tests of the internal consistency of the forecasts across two or more horizons. For example, forecasts of an artificially-generated AR(p) process, independent of the actual series but constructed in a theoretically optimal fashion, would not be identified as suboptimal by this test.²

Example 2: Consider a scenario where all forecasts are contaminated with noise (due, e.g., to estimation error) that is increasing in the forecast horizon:

$$\begin{aligned} \tilde{Y}_{t|t-h_L} &= \hat{Y}_{t|t-h_L}^* + u_{t-h_L}, \quad u_{t-h_L} \perp \hat{Y}_{t|t-h_L} \\ \tilde{Y}_{t|t-h_S} &= \hat{Y}_{t|t-h_S}^* + u_{t-h_S}, \quad u_{t-h_S} \perp (\hat{Y}_{t|t-h_L}, \hat{Y}_{t|t-h_S}, u_{t-h_L}) \\ V[u_{t-h_L}] &> V[u_{t-h_S}]. \end{aligned}$$

Define the forecast revision from time $t - h_L$ to $t - h_S$ as $\eta_{t|h_S, h_L} \equiv \hat{Y}_{t|t-h_S}^* - \hat{Y}_{t|t-h_L}^*$. Note that by forecast optimality we have $Cov[\hat{Y}_{t|t-h_L}^*, \eta_{t|h_S, h_L}] = 0$, and so:

$$\begin{aligned} V[\tilde{Y}_{t|t-h_S}] - V[\tilde{Y}_{t|t-h_L}] &= V[\hat{Y}_{t|t-h_S}^*] + V[u_{t-h_S}] - V[\hat{Y}_{t|t-h_L}^*] - V[u_{t-h_L}] \\ &= V[\hat{Y}_{t|t-h_L}^* + \eta_{t|h_S, h_L}] + V[u_{t-h_S}] - V[\hat{Y}_{t|t-h_L}^*] - V[u_{t-h_L}] \\ &= V[\eta_{t|h_S, h_L}] + V[u_{t-h_S}] - V[u_{t-h_L}] \\ &< 0 \quad \text{if } V[u_{t-h_L}] > V[\eta_{t|h_S, h_L}] + V[u_{t-h_S}]. \end{aligned}$$

Hence, if the contaminating noise in the long-horizon forecast is greater than the sum of the variance of the optimal forecast revision and the variance of the short-horizon noise, the long-horizon forecast will have greater variance than the short-horizon forecast, and a test based on (6) should detect this.

Note that the violation of forecast optimality discussed in Example 1, with the short-horizon forecast generated as the long-horizon forecast plus some independent noise, would *not* be detected as sub-optimal by a test of the monotonicity of the mean-squared forecast. In this case the short-horizon forecast would indeed be more volatile than the long-horizon forecast, consistent with

²For tests of internal consistency across point forecasts and density forecasts, see Clements (2009).

optimality, and this test would not be able to detect that the source of this increased variation was simply uninformative noise.

2.5 Monotonicity of covariance between the forecast and target variable

An implication of the weakly decreasing forecast variance property established in Corollary 2 is that the covariance of the forecasts with the target variable should be *decreasing* in the forecast horizon. To see this, note that

$$\text{Cov} \left[\hat{Y}_{t|t-h}^*, Y_t \right] = \text{Cov} \left[\hat{Y}_{t|t-h}^*, \hat{Y}_{t|t-h}^* + e_{t|t-h}^* \right] = V \left[\hat{Y}_{t|t-h}^* \right].$$

Thus we obtain the following:

Corollary 3 *Under the assumptions of Theorem 1 and S1, we obtain*

$$\text{Cov} \left[\hat{Y}_{t|t-h_S}^*, Y_t \right] \geq \text{Cov} \left[\hat{Y}_{t|t-h_L}^*, Y_t \right] \quad \text{for any } h_S < h_L.$$

Further, since $E \left[\hat{Y}_{t|t-h}^* \right] = E \left[Y_t \right]$, we also obtain:

$$E \left[\hat{Y}_{t|t-h_S}^* Y_t \right] \geq E \left[\hat{Y}_{t|t-h_L}^* Y_t \right] \quad \text{for any } h_S < h_L.$$

As for the above cases, this implication can again be tested using Wolak's (1989) approach by defining the vector $\Delta^c \equiv [\Delta_2^c, \dots, \Delta_H^c]'$, where $\Delta_j^c \equiv E \left[\hat{Y}_{t|t-h_j}^* Y_t \right] - E \left[\hat{Y}_{t|t-h_{j-1}}^* Y_t \right]$ and testing the null hypothesis:

$$H_0 \quad : \quad \Delta^c \leq \mathbf{0}, \tag{8}$$

$$\text{vs. } H_1 \quad : \quad \Delta^c \in \mathbb{R}^{H-1}.$$

2.6 Monotonicity of mean squared forecast revisions

Monotonicity of mean squared forecasts also implies a monotonicity result for the mean squared forecast *revisions*. Consider the following decomposition of the short-horizon forecast into the long-horizon forecast plus the sum of forecast revisions:

$$\begin{aligned} \hat{Y}_{t|t-h_1}^* &= \hat{Y}_{t|t-h_H}^* + \left(\hat{Y}_{t|t-h_{H-1}}^* - \hat{Y}_{t|t-h_H}^* \right) + \dots + \left(\hat{Y}_{t|t-h_1}^* - \hat{Y}_{t|t-h_2}^* \right) \\ &\equiv \hat{Y}_{t|t-h_H}^* + \sum_{j=1}^{H-1} \eta_{t|h_j, h_{j+1}}, \end{aligned} \tag{9}$$

where $\eta_{t|h_S, h_L} \equiv \hat{Y}_{t|t-h_S}^* - \hat{Y}_{t|t-h_L}^*$ for $h_S < h_L$. Under optimality, $E_{t-h_L} [\eta_{t|h_S, h_L}] = 0$ for all $h_S < h_L$, so $Cov [\hat{Y}_{t|t-h_L}^*, \eta_{t|h_S, h_L}] = 0$, and

$$\begin{aligned} V [\hat{Y}_{t|t-h_1}^* - \hat{Y}_{t|t-h_H}^*] &= V \left[\sum_{j=1}^{H-1} \eta_{t|h_j, h_{j+1}} \right] = \sum_{j=1}^{H-1} V [\eta_{t|h_j, h_{j+1}}], \\ V [\hat{Y}_{t|t-h_1}^* - \hat{Y}_{t|t-h_{H-1}}^*] &= \sum_{j=1}^{H-2} V [\eta_{t|h_j, h_{j+1}}] \leq V [\hat{Y}_{t|t-h_1}^* - \hat{Y}_{t|t-h_H}^*] \end{aligned}$$

More generally, the following corollary to Theorem 1 holds:

Corollary 4 Denote the forecast revision between two dates as $\eta_{t|h_S, h_L} \equiv \hat{Y}_{t|t-h_S}^* - \hat{Y}_{t|t-h_L}^*$ for any $h_S < h_L$. Under the assumptions of Theorem 1 and S1, we have

$$V [\eta_{t|h_S, h_L}] \geq V [\eta_{t|h_S, h_M}] \quad \text{for any } h_S < h_M < h_L.$$

Further, since $E [\eta_{t|h_S, h_L}] = 0$, we also obtain:

$$E [\eta_{t|h_S, h_L}^2] \geq E [\eta_{t|h_S, h_M}^2] \quad \text{for any } h_S < h_M < h_L. \quad (10)$$

Considering the forecast revisions between each horizon and the shortest horizon, this implies that

$$V [\eta_{t|h_1, h_2}] \leq V [\eta_{t|h_1, h_3}] \leq \dots \leq V [\eta_{t|h_1, h_H}]. \quad (11)$$

Again Wolak's (1987, 1989) testing framework can be applied here: Define the vector of mean-squared forecast revisions $\Delta^\eta \equiv [\Delta_3^\eta, \dots, \Delta_H^\eta]'$, where $\Delta_j^\eta \equiv E [\eta_{t|h_1, h_j}^2] - E [\eta_{t|h_1, h_{j-1}}^2] = E \left[\left(\hat{Y}_{t|t-h_1}^* - \hat{Y}_{t|t-h_j}^* \right)^2 \right] - E \left[\left(\hat{Y}_{t|t-h_1}^* - \hat{Y}_{t|t-h_{j-1}}^* \right)^2 \right]$. Then we can test the null hypothesis that the differences in mean-squared forecast revisions are weakly positive for all forecast horizons:

$$\begin{aligned} H_0 &: \Delta^\eta \geq \mathbf{0}, \\ \text{vs. } H_1 &: \Delta^\eta \in \mathbb{R}^{H-2}. \end{aligned} \quad (12)$$

Example 3: Consider forecasts with either “sticky” updating or, conversely, “overshooting”:

$$\hat{Y}_{t|t-h} = \gamma \hat{Y}_{t|t-h}^* + (1 - \gamma) \hat{Y}_{t|t-h-1}^*, \quad \text{for } h = 1, 2, \dots, H.$$

“Sticky” forecasts correspond to $\gamma \in [0, 1)$, while “overshooting” occurs when $\gamma > 1$. Moreover, suppose the underlying data generating process is an AR(1), $Y_t = \phi Y_{t-1} + \varepsilon_t$, $|\phi| < 1$, so $\hat{Y}_{t|t-h}^* =$

$\phi^h Y_{t-h}$. Then we have

$$\begin{aligned}
\eta_{t|t-h,t-h-1} &= \gamma(\hat{Y}_{t|t-h}^* - \hat{Y}_{t|t-h-1}^*) + (1-\gamma)(\hat{Y}_{t|t-h-1}^* - \hat{Y}_{t|t-h-2}^*) \\
&= \gamma(\phi^h Y_{t-h} - \phi^{h+1} Y_{t-h-1}) + (1-\gamma)(\phi^{h+1} Y_{t-h-1} - \phi^{h+2} Y_{t-h-2}) \\
&= \gamma\phi^h \varepsilon_{t-h} + (1-\gamma)\phi^{h+1} \varepsilon_{t-h-1}.
\end{aligned}$$

It follows that the variances of the one- and two-period forecast revisions are

$$\begin{aligned}
V(\eta_{t|1,2}) &= (\gamma^2\phi^2 + (1-\gamma)^2\phi^4) \sigma_\varepsilon^2, \\
V(\eta_{t|1,3}) &= (\gamma^2\phi^2 + \phi^4 + (1-\gamma)^2\phi^6) \sigma_\varepsilon^2.
\end{aligned}$$

We can then have a violation of the inequality in (11), if

$$(1-\gamma)^2 > 1 + (1-\gamma)^2\phi^2.$$

There is clearly no violation if $\gamma = 1$ (full optimality) or if γ is close to one. However, if γ is far from one, representing either very sticky forecasts (e.g., $\gamma = 0.5$) or overshooting (e.g., $\gamma = 1.5$).

2.7 Bounds on covariances of forecast revisions

Combining the inequalities contained in the above corollaries, it turns out that we can place an upper bound on the variance of the forecast revision, as a function of the covariance of the revision with the target variable. The intuition behind this bound is simple: if little relevant information arrives between the updating points, then the variance of the forecast revisions must be low.

Corollary 5 *Denote the forecast revision between two dates as $\eta_{t|h_S, h_L} \equiv \hat{Y}_{t|t-h_S}^* - \hat{Y}_{t|t-h_L}^*$ for any $h_S < h_L$. Under the assumptions of Theorem 1 and S1, we have*

$$V[\eta_{t|h_S, h_L}] \leq 2\text{Cov}[Y_t, \eta_{t|h_S, h_L}] \quad \text{for any } h_S < h_L.$$

Further, since $E[\eta_{t|h_S, h_L}] = 0$, we also obtain:

$$E[\eta_{t|h_S, h_L}^2] \leq 2E[Y_t \eta_{t|h_S, h_L}] \quad \text{for any } h_S < h_L. \quad (13)$$

Note also that this result implies (as one would expect) that the covariance between the target variable and the forecast revision must be positive; when forecasts are updated to reflect new information, the change in the forecast should be positively correlated with the target variable.

The above bound can be tested by re-writing it as

$$E \left[2Y_t \eta_{t|h_S, h_L} - \eta_{t|h_S, h_L}^2 \right] \geq 0. \quad (14)$$

and forming the vector $\Delta^b \equiv [\Delta_2^b, \dots, \Delta_H^b]'$, where $\Delta_j^b \equiv E \left[2Y_t \eta_{t|h_j, h_{j-1}} - \eta_{t|h_j, h_{j-1}}^2 \right]$, for $j = 2, \dots, H$ and then testing the null hypothesis that this variable is weakly positive for all forecast horizons

$$\begin{aligned} H_0 &: \Delta^b \geq \mathbf{0} \\ \text{vs. } H_1 &: \Delta^b \in \mathbb{R}^{H-1}. \end{aligned}$$

Example 1, continued: Consider once again the case where the short-horizon forecast is equal to the long-horizon forecast plus noise:

$$\tilde{Y}_{t|h_S} = \tilde{Y}_{t|h_L} + u_{t-h_S}, \quad u_{t-h_S} \perp \tilde{Y}_{t|h_L}.$$

The difference between the variance of the forecast revision and twice the covariance of the revision with the target variable (which is negative under forecast optimality) now equals the variance of the noise:

$$\begin{aligned} V \left[\tilde{\eta}_{t|h_S, h_L} \right] - 2Cov \left[Y_t, \tilde{\eta}_{t|h_S, h_L} \right] &= V \left[\tilde{Y}_{t|h_L} + u_{t-h_S} - \tilde{Y}_{t|h_L} \right] - 2Cov \left[Y_t, \tilde{Y}_{t|h_L} + u_{t-h_S} - \tilde{Y}_{t|h_L} \right] \\ &= V \left[u_{t-h_S} \right] > 0. \end{aligned}$$

Here the bound is violated: The extra noise in the short-horizon forecast contributes to the variance of the forecast revision without increasing the covariance of the revision with the target variable.

Example 2, continued: Consider again the case where all forecasts are contaminated with noise, $\tilde{Y}_{t|h} = \hat{Y}_{t|h} + u_{t|h}$, whose variance is increasing in the forecast horizon, $V \left[u_{t|h_L} \right] > V \left[u_{t|h_S} \right]$. In this case we find:

$$\begin{aligned} V \left[\tilde{\eta}_{t|h_S, h_L} \right] - 2Cov \left[Y_t, \tilde{\eta}_{t|h_S, h_L} \right] &= V \left[\eta_{t|h_S, h_L} + u_{t|h_S} - u_{t|h_L} \right] - 2Cov \left[Y_t, \eta_{t|h_S, h_L} + u_{t|h_S} - u_{t|h_L} \right] \\ &= \left\{ V \left[\eta_{t|h_S, h_L} \right] - 2Cov \left[Y_t, \eta_{t|h_S, h_L} \right] \right\} + V \left[u_{t|h_S} \right] + V \left[u_{t|h_L} \right]. \end{aligned}$$

Under forecast optimality we know that the term in braces is negative, but if the sum of the short-horizon and long-horizon noise is greater in absolute value than the term in braces, we will observe a violation of the bound, and a test of this bound can be used to reject forecast optimality.

2.8 Variance bounds tests without data on the target variable

The “real time” macroeconomics literature has demonstrated the presence of large and prevalent measurement errors affecting a variety of macroeconomic variables, see Croushore (2006), Croushore and Stark (2001), Faust, Rogers and Wright (2005), and Corradi, Fernandez and Swanson (2009). In such situations it is useful to have tests that do not require data on the target variable. Corollaries 2 and 4 presented two testable implications of forecast optimality that do not require data on the target variable, and in this section we present further tests of multi-horizon forecast optimality that can be employed when data on the target variable is not available or is not reliable.

The tests in this section exploit the fact that, under the null of forecast optimality, the short-horizon forecast can be taken as a proxy for the target variable, from the stand-point of longer-horizon forecasts, in the sense that the inequality results presented above all hold when the short-horizon forecast is used in place of the target variable. Importantly, unlike standard cases, the proxy in this case is *smoother* rather than noisier than the actual variable. This turns out to have beneficial implications for the finite-sample performance of these tests when the measurement error is sizeable or the predictive R^2 of the forecasting model is low.

The result that corresponds to that in Corollary 1 is presented in Corollary 4. The corresponding results for Corollaries 3 and 5 are presented below:

Corollary 6 *Under the assumptions of Theorem 1 and S1 we obtain:*

(a)

$$\begin{aligned} \text{Cov} \left[\hat{Y}_{t|t-h_M}^*, \hat{Y}_{t|t-h_S}^* \right] &\geq \text{Cov} \left[\hat{Y}_{t|t-h_L}^*, \hat{Y}_{t|t-h_S}^* \right], \text{ and} \\ E \left[\hat{Y}_{t|t-h_M}^* \hat{Y}_{t|t-h_S}^* \right] &\geq E \left[\hat{Y}_{t|t-h_L}^* \hat{Y}_{t|t-h_S}^* \right] \text{ for any } h_S < h_M < h_L \end{aligned} \quad (15)$$

(b) *Denote the forecast revision between two dates as $\eta_{t|h,k} \equiv \hat{Y}_{t|t-h}^* - \hat{Y}_{t|t-k}^*$ for any $h < k$.*

Then

$$\begin{aligned} V \left[\eta_{t|h_M, h_L} \right] &\leq 2 \text{Cov} \left[\hat{Y}_{t|t-h_S}^*, \eta_{t|h_M, h_L} \right], \text{ and} \\ E \left[\eta_{t|h_M, h_L}^2 \right] &\leq 2 E \left[\hat{Y}_{t|t-h_S}^* \eta_{t|h_M, h_L} \right] \text{ for any } h_S < h_M < h_L. \end{aligned} \quad (16)$$

As for Corollaries 3 and 5, an inevitable side-effect of testing forecast optimality without using data on the target variable is that such a test only examines the internal consistency of the forecasts across the different horizons; an internally consistent set of forecasts that are not optimal for a given target variable will not be detected using such tests.

2.9 Illustration for an AR(1) process

This section illustrates the above results for the special case of an AR(1) process. Let:

$$Y_t = \phi Y_{t-1} + \varepsilon_t, \quad |\phi| < 1, \quad (17)$$

where $\varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$, so $\sigma_y^2 = \sigma_\varepsilon^2 / (1 - \phi^2)$. Rewriting this as

$$Y_t = \phi^h Y_{t-h} + \sum_{i=0}^{h-1} \phi^i \varepsilon_{t-i},$$

we have $\hat{Y}_{t|t-h}^* = \phi^h Y_{t-h}$, and so $e_{t|t-h}^* = \sum_{i=0}^{h-1} \phi^i \varepsilon_{t-i}$. From this it follows that, consistent with Corollary 1,

$$V[e_{t|t-h}^*] = \sigma_\varepsilon^2 \left(\frac{1 - \phi^{2h}}{1 - \phi^2} \right) \leq \sigma_\varepsilon^2 \left(\frac{1 - \phi^{2(h+1)}}{1 - \phi^2} \right) = V[e_{t|t-h-1}^*].$$

Moreover, consistent with Corollary 2 the variance of the forecast is increasing in h :

$$V[\hat{Y}_{t|t-h}^*] = \phi^{2h} \sigma_y^2 \geq \phi^{2(h+1)} \sigma_y^2 = V[\hat{Y}_{t|t-h-1}^*].$$

The covariance between the outcome and the h -period forecast is

$$Cov[Y_t, \hat{Y}_{t|t-h}^*] = Cov\left[\phi^h Y_{t-h} + \sum_{i=0}^{h-1} \phi^i \varepsilon_{t-i}, \phi^h Y_{t-h}\right] = \phi^{2h} \sigma_y^2,$$

which is decreasing in h , consistent with Corollary 3. Also, noting that $\hat{Y}_{t|t-h_S}^* = \hat{Y}_{t|t-h_L}^* + \sum_{i=h_S}^{h_L-1} \phi^i \varepsilon_{t-i}$, the forecast revision can be written as $\eta_{t|h_S, h_L} = \sum_{i=h_S}^{h_L-1} \phi^i \varepsilon_{t-i}$, and so

$$V[\eta_{t|h_S, h_L}] = \sigma_\varepsilon^2 \phi^{2h_S} \left(\frac{1 - \phi^{2(h_L - h_S)}}{1 - \phi^2} \right),$$

which is increasing in $h_L - h_S$, consistent with Corollary 4. Consistent with Corollary 5, the variance of the revision is bounded by twice the covariance of the actual value and the revision:

$$2Cov[Y_t, \eta_{t|h_S, h_L}] = 2V\left[\sum_{i=h_S}^{h_L-1} \phi^i \varepsilon_{t-i}\right] > V\left[\sum_{i=h_S}^{h_L-1} \phi^i \varepsilon_{t-i}\right] = \eta_{t|h_S, h_L}.$$

The implications of forecast rationality presented in Corollary 6 for this AR(1) example are:

$$\begin{aligned} Cov[\hat{Y}_{t|t-h_M}^*, \hat{Y}_{t|t-h_S}^*] &= Cov[\hat{Y}_{t|t-h_M}^*, \hat{Y}_{t|t-h_M}^* + \eta_{t|h_S, h_M}] \\ &= Cov\left[\phi^{h_M} Y_{t-h_M}, \phi^{h_M} Y_{t-h_M} + \sum_{i=h_S}^{h_M-1} \phi^i \varepsilon_{t-i}\right] \\ &= \phi^{2h_M} V[Y_{t-h_M}] = \phi^{2h_M} \frac{\sigma_\varepsilon^2}{1 - \phi^2} \geq \phi^{2h_L} \frac{\sigma_\varepsilon^2}{1 - \phi^2} = Cov[\hat{Y}_{t|t-h_L}^*, \hat{Y}_{t|t-h_S}^*] \end{aligned}$$

and

$$\begin{aligned}
Cov \left[\hat{Y}_{t|h_S}^*, \eta_{t|h_M, h_L} \right] &= Cov \left[\phi^{h_S} Y_{t-h_S}, \sum_{i=h_M}^{h_L-1} \phi^i \varepsilon_{t-i} \right] \\
&= Cov \left[\hat{Y}_{t|h_S}^* + \sum_{i=h_S}^{h_M-1} \phi^i \varepsilon_{t-i} + \sum_{i=h_M}^{h_L-1} \phi^i \varepsilon_{t-i}, \sum_{i=h_M}^{h_L-1} \phi^i \varepsilon_{t-i} \right] \\
&= V \left[\sum_{i=h_M}^{h_L-1} \phi^i \varepsilon_{t-i} \right] = \sigma_\varepsilon^2 \sum_{i=h_M}^{h_L-1} \phi^{2i} = \sigma_\varepsilon^2 \phi^{2h_M} \frac{1 - \phi^{2(h_L-h_M)}}{1 - \phi^2} \\
\text{while } V \left[\eta_{t|h_M, h_L} \right] &= V \left[\sum_{i=h_M}^{h_L-1} \phi^i \varepsilon_{t-i} \right] = \sigma_\varepsilon^2 \sum_{i=h_M}^{h_L-1} \phi^{2i} = \sigma_\varepsilon^2 \phi^{2h_M} \frac{1 - \phi^{2(h_L-h_M)}}{1 - \phi^2} \\
&\leq 2Cov \left[\hat{Y}_{t|h_S}^*, \eta_{t|h_M, h_L} \right]
\end{aligned}$$

3 Regression Tests of Forecast Rationality

Conventional Mincer-Zarnowitz (MZ) regression tests form a natural benchmark against which the performance of our new optimality tests can be compared, both because they are in widespread use and because they are easy to implement. Such regressions test directly if forecast errors are orthogonal to variables contained in the forecaster's information set. For a single forecast horizon, h , the standard Mincer-Zarnowitz (MZ) regression takes the form:

$$Y_t = \alpha_h + \beta_h \hat{Y}_{t|h} + u_{t|h}, \quad (18)$$

while forecast optimality can be tested through an implication of optimality that we summarize in the following corollary to Theorem 1:

Corollary 7 *Under the assumptions of Theorem 1 and S1, the population values of the parameters in the Mincer-Zarnowitz regression in equation (18) satisfy*

$$H_0^h : \alpha_h = 0 \cap \beta_h = 1, \text{ for each horizon } h.$$

The MZ regression in (18) is usually applied separately to each forecast horizon. A simultaneous test of optimality across all horizons requires developing a different approach. We next present two standard ways of combining these results.

3.1 Bonferroni bounds on MZ regressions

One approach, adopted in Capistrán (2007), is to run MZ regressions (18) for each horizon, $h = h_1, \dots, h_H$. For each forecast horizon, h , we can obtain the p -value from a chi-squared test with two degrees of freedom. A Bonferroni bound is then used to obtain a joint test. In particular, we reject forecast optimality if the minimum p -value across all H tests is less than the desired size divided by H , α/H . This approach is often quite conservative.

3.2 Vector MZ tests

An alternative to the Bonferroni-bounds approach is to stack the MZ equations for each horizon and estimate the model as a system:

$$\begin{bmatrix} Y_{t+h_1} \\ Y_{t+h_2} \\ Y_{t+h_3} \\ \vdots \\ Y_{t+h_H} \end{bmatrix} = \begin{bmatrix} \alpha_{h_1} \\ \alpha_{h_2} \\ \alpha_{h_3} \\ \vdots \\ \alpha_{h_H} \end{bmatrix} + \begin{bmatrix} \beta_{h_1} & 0 & 0 & \cdots & 0 \\ 0 & \beta_{h_2} & 0 & \cdots & 0 \\ 0 & 0 & \beta_{h_3} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \beta_{h_H} \end{bmatrix} \begin{bmatrix} \hat{Y}_{t+h_1|t} \\ \hat{Y}_{t+h_2|t} \\ \hat{Y}_{t+h_3|t} \\ \vdots \\ \hat{Y}_{t+h_H|t} \end{bmatrix} + \begin{bmatrix} e_{t+h_1}^1 \\ e_{t+h_2}^2 \\ e_{t+h_3}^3 \\ \vdots \\ e_{t+h_H}^H \end{bmatrix}. \quad (19)$$

The relevant hypothesis is now

$$H_0 : \alpha_{h_1} = \dots = \alpha_{h_H} = 0 \cap \beta_{h_1} = \dots = \beta_{h_H} = 1 \quad (20)$$

$$H_1 : \alpha_{h_1} \neq 0 \cup \dots \cup \alpha_{h_H} \neq 0 \cup \beta_{h_1} \neq 1 \cup \dots \cup \beta_{h_H} \neq 1.$$

For $h > 1$, the residuals in (19) will, even under the null of optimality, exhibit autocorrelation and will typically also exhibit cross-autocorrelation, so a HAC estimator of the standard errors is required.

3.3 Univariate Optimal Revision Regression

We next propose a new approach to test optimality that utilizes the complete set of forecasts in the context of univariate regressions. The approach is to estimate a *univariate* regression of the target variable on the longest-horizon forecast, $\hat{Y}_{t|t-h_H}$, and all the intermediate forecast revisions, $\eta_{t|h_1, h_2}, \dots, \eta_{t|h_{H-1}, h_H}$. To derive this test, notice that we can represent a short-horizon forecast as

a function of a long-horizon forecast and the intermediate forecast revisions:

$$\hat{Y}_{t|t-h_1} \equiv \hat{Y}_{t|t-h_H} + \sum_{j=1}^{H-1} \eta_{t|h_j, h_{j+1}}.$$

Rather than regressing the outcome variable on the one-period forecast, we proposed the following “optimal revision” regression:

$$Y_t = \alpha + \beta_H \hat{Y}_{t|t-h_H} + \sum_{j=1}^{H-1} \beta_j \eta_{t|h_j, h_{j+1}} + u_t. \quad (21)$$

Corollary 8 *Under the assumptions of Theorem 1 and S1, the population values of the parameters in the optimal revision regression in equation (21) satisfy*

$$H_0 : \alpha = 0 \cap \beta_1 = \dots = \beta_H = 1.$$

The regression in equation (21) can be re-written as the target variable on all of the forecasts, from h_1 to h_H , and the parameter restrictions given in Corollary 8 are then that the intercept is zero, the coefficient on the short-horizon forecast is one, and the coefficients on all longer-horizon forecasts are zero.

This univariate regression tests both that agents optimally and consistently revise their forecasts at the interim points between the longest and shortest forecast horizons and also that the long-run forecast is unbiased. Hence it generalizes the conventional Mincer-Zarnowitz regression (18) which only considers a single horizon.

3.4 Regression tests without the target variable

All three of the above regression-based tests above can be applied with the short-horizon forecast used in place of the target variable. That is, we can undertake a Mincer-Zarnowitz regression of the short-horizon forecast on a long-horizon forecast

$$\hat{Y}_{t|t-h_1} = \tilde{\alpha}_j + \tilde{\beta}_j \hat{Y}_{t|t-h_j} + v_{t|t-h_j} \quad \text{for all } h_j > h_1. \quad (22)$$

Similarly, we get a vector MZ test that uses forecasts as target variables:

$$\begin{bmatrix} \hat{Y}_{t+h_2|t+h_2-1} \\ \vdots \\ \hat{Y}_{t+h_H|t+h_H-1} \end{bmatrix} = \begin{bmatrix} \tilde{\alpha}_2 \\ \vdots \\ \tilde{\alpha}_H \end{bmatrix} + \begin{bmatrix} \tilde{\beta}_2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \tilde{\beta}_H \end{bmatrix} \begin{bmatrix} \hat{Y}_{t+h_2|t} \\ \vdots \\ \hat{Y}_{t+h_H|t} \end{bmatrix} + \begin{bmatrix} v_{t|t-h_2} \\ \vdots \\ v_{t|t-h_H} \end{bmatrix}. \quad (23)$$

And finally we can estimate a version of the optimal revision regression:

$$\hat{Y}_{t|t-h_1} = \tilde{\alpha} + \tilde{\beta}_H \hat{Y}_{t|t-h_H} + \sum_{j=2}^{H-1} \tilde{\beta}_j \eta_{t|h_j, h_{j+1}} + u_t, \quad (24)$$

The parameter restrictions implied by forecast optimality are the same as in the standard cases, and are presented in the following corollary:

Corollary 9 *Under the assumptions of Theorem 1 and S1, the population values of the parameters in (a) Mincer-Zarnowitz regression by proxy in equation (22) satisfy*

$$H_0^h : \tilde{\alpha}_h = 0 \cap \tilde{\beta}_h = 1, \text{ for each horizon } h > h_1,$$

and (b) the population values of the parameters in the optimal revision regression by proxy, in equation (24) satisfy

$$H_0 : \tilde{\alpha} = 0 \cap \tilde{\beta}_2 = \dots = \tilde{\beta}_H = 1.$$

This result exploits the fact that under optimality (and squared error loss) each forecast can be considered a conditionally unbiased proxy for the (unobservable) target variable, where the conditioning is on the information set available at the time the forecast is made. That is, if $\hat{Y}_{t|t-h_S} = E_{t-h_S}[Y_t]$ for all h_S , then $E_{t-h_L}[\hat{Y}_{t|t-h_S}] = E_{t-h_L}[Y_t]$ for any $h_L > h_S$, and so the short-horizon forecast is a conditionally unbiased proxy for the realization. If forecasts from multiple horizons are available, then we can treat the short-horizon forecast as a proxy for the actual variable, and use it to “test the optimality” of the long-horizon forecast. In fact, this regression tests the internal consistency of the two forecasts, and thus tests an implication of the null that *both* forecasts are rational.

4 Stationarity and Tests of Forecast Optimality (*incomplete*)

The literature on forecast evaluation conventionally assumes that the underlying data generating process is covariance stationary. Under this assumption, the Wold decomposition applies and so

$$Y_t = f(t, \theta) + Y_0 + \tilde{Y}_t,$$

where $f(t, \theta)$ captures deterministic parts (e.g. seasonality or trends); Y_0 represents the initial condition and \tilde{Y}_t is the covariance stationary component which has the Wold representation

$$\tilde{Y}_t = \sum_{i=0}^t \theta_i \varepsilon_{t-i}, \quad (25)$$

where $\varepsilon_{t-i} \sim WN(0, 1)$ is serially uncorrelated white noise and $\lim_{t \rightarrow \infty} \sum_{i=0}^t \theta_i^2 < \infty$. Forecast analysis often focuses on the covariance stationary component, \tilde{Y}_t . If the underlying data is non-stationary, typically stationarity is recovered by appropriately first- or second-differencing the data.

To see the role played by the covariance stationarity assumption, let $\hat{Y}_{t+h|t-j}^* = \arg \min_{\hat{Y}_{t+h|t-j}} E_{t-j}[(Y_{t+h} - \hat{Y}_{t+h|t-j})^2]$. By optimality, we must have

$$E_t[(Y_{t+h} - \hat{Y}_{t+h|t-j}^*)^2] \geq E_t[(Y_{t+h} - \hat{Y}_{t+h|t}^*)^2] \quad \text{for } j \geq 1. \quad (26)$$

Then, by the law of iterated expectations,

$$E[(Y_{t+h} - \hat{Y}_{t+h|t-j}^*)^2] \geq E[(Y_{t+h} - \hat{Y}_{t+h|t}^*)^2] \quad \text{for } j \geq 1. \quad (27)$$

This result compares the variance of the error in predicting the outcome at time $t+h$ given information at time t against the prediction error given information at an earlier date, $t-j$. Usually, however, forecast comparisons are based on forecasts made at the same date, t , and hence conditional on the same information set, \mathcal{F}_t , but for different forecast horizons, corresponding to predicting Y_{t+h+j} and Y_{t+h} given \mathcal{F}_t . Provided that $(Y_{t+h} - \hat{Y}_{t+h|t-j}^*)$ is covariance stationary, it follows from (27) that

$$E[(Y_{t+h+j} - \hat{Y}_{t+h+j|t}^*)^2] \geq E[(Y_{t+h} - \hat{Y}_{t+h|t}^*)^2] \quad \text{for } j \geq 1. \quad (28)$$

The covariance stationarity assumption is clearly important here. (28) does not follow from (27) if, say, there is a deterministic reduction in the variance of Y between periods $t+h$ and $t+h+j$. Suppose for example that

$$Y_\tau = \begin{cases} \mu + \sigma \varepsilon_\tau & \text{for } \tau \leq t+h \\ \mu + \frac{\sigma}{2} \varepsilon_\tau & \text{for } \tau > t+h \end{cases}, \quad (29)$$

where ε_τ is zero-mean white noise. This could be a stylized example of the “Great Moderation”.

Clearly (28) is now violated as $\hat{Y}_{t+h+j|t}^* = \hat{Y}_{t+h|t}^* = \mu$, and so³

$$E[(Y_{t+h+j} - \hat{Y}_{t+h+j|t}^*)^2] = \frac{\sigma^2}{4} < \sigma^2 = E[(Y_{t+h} - \hat{Y}_{t+h|t}^*)^2] \quad \text{for } j \geq 1. \quad (30)$$

For example, in the case of the Great Moderation, which is believed to have occurred around 1984, a one-year-ahead forecast made in 1982 (i.e. for GDP growth in 1983, while volatility was still high) could well be associated with greater errors than, say, a three-year-ahead forecast (i.e. for GDP growth in 1985, after volatility has come down).

³Notice here that the expectation, $E[\cdot]$, is taken under the assumption that we know the break in the variance since this is assumed to be deterministic.

4.1 Fixed event forecasts

Under covariance stationarity, studying the precision of a sequence of forecasts $\hat{Y}_{t|t-h}$ is equivalent to comparing the precision of $\hat{Y}_{t+h|t}$ for different values of h . However, this equivalence need not hold when the predicted variable, Y_t , is not covariance stationary. One way to deal with non-stationarities such as the break in the variance in (29) is to hold the forecast ‘event’ fixed, while varying the time to the event, h . In this case the forecast optimality test gets based on (27) rather than (28). Forecasts where the target date, t , is kept fixed, while the forecast horizon varies are commonly called fixed-event forecasts, see Clements (1997) and Nordhaus (1987).

To see how this works, notice that, by forecast optimality,

$$E_{t-h_S}[(Y_t - \hat{Y}_{t|t-h_L}^*)^2] \geq E_{t-h_S}[(Y_t - \hat{Y}_{t|t-h_S}^*)^2] \quad \text{for } h_L \geq h_S. \quad (31)$$

Moreover, by the law of iterated expectations,

$$E[(Y_t - \hat{Y}_{t|t-h_L}^*)^2] \geq E[(Y_t - \hat{Y}_{t|t-h_S}^*)^2] \quad \text{for } h_L \geq h_S. \quad (32)$$

This result is quite robust. For example, with a break in the variance, (29), we have $\hat{Y}_{t|t-h_L}^* = \hat{Y}_{t|t-h_S}^* = \mu$, and

$$E[(Y_\tau - \hat{Y}_{\tau|t-h_L}^*)^2] = E[(Y_\tau - \hat{Y}_{\tau|t-h_S}^*)^2] = \begin{cases} \sigma^2 & \text{for } \tau \leq t+h \\ \sigma^2/4 & \text{for } \tau > t+h \end{cases}.$$

As a second example, suppose we let the mean of a time-series be subject to a probabilistic break that only is known once it has happened. To this end, define an absorbing state process, $s_\tau \in \mathcal{F}_\tau$, such that $s_0 = 0$ and $\Pr(s_\tau = 0 | s_{\tau-1} = 0) = \pi$ for all τ . Consider the following process:

$$y_\tau = \mu + s_\tau \Delta_\mu + (\sigma + s_\tau \Delta_\sigma) \varepsilon_\tau, \quad \varepsilon_\tau \sim (0, 1).$$

Suppose we condition on $s_{t-h} = 0$, $s_{t-h+1} = 1$, so the permanent break happens at time $t-h+1$. Then

$$\hat{Y}_{t|t-j}^* = \begin{cases} \mu + \pi \Delta_\mu & \text{for } j \geq h \\ \mu + \Delta_\mu & \text{for } j \leq h-1 \end{cases},$$

and so we have the expected loss

$$E_{t-j}[(Y_t - \hat{Y}_{t|t-j}^*)^2] = \begin{cases} (1-\pi)^2 \Delta_\mu^2 + (\sigma + \Delta_\sigma)^2 & \text{for } j \geq h \\ (\sigma + \Delta_\sigma)^2 & \text{for } j \leq h-1 \end{cases}.$$

Once again, monotonicity continues to hold for the fixed-event forecasts, i.e. for $h_L > h_S$ and for all t :

$$E[(Y_t - \hat{Y}_{t|t-h_L}^*)^2] \geq E[(Y_t - \hat{Y}_{t|t-h_S}^*)^2].$$

4.2 A general class of non-stationary processes

Provided that a fixed-event setup is used, we next show that the variance bound results pertain to a more general class of stochastic processes that do not require covariance stationarity.

Assumption S2: The target variable, Y_t , is generated by

$$Y_t = f(t, \theta) + \sum_{i=0}^{\infty} \theta_{it} \varepsilon_{t-i}. \quad (33)$$

where $f(t, \theta)$ captures deterministic parts (e.g. seasonality or trends), Y_0 represents the initial condition, $\varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$ is serially uncorrelated mean-zero white noise, and, for all t , θ_{it} is a sequence of deterministic coefficients such that $\sum_{i=0}^{\infty} \theta_{it}^2 < \infty$ for all t .

Assumption S1 (covariance stationarity) is sufficient for S2, and further implies that the coefficients in the representation are not functions of t , i.e., $\theta_{it} = \theta_i \forall t$. However Assumption S2 allows for deviations from covariance stationarity that can be modeled via deterministic changes in the usual Wold decomposition weights, θ_i . For example, it may be that, due to a change in economic policy or the economic regime, the impulse response function changes after a certain date. For the example in (29), we get

$$\theta_{0,\tau} = \begin{cases} \sigma & \text{for } \tau \leq t+h \\ \frac{\sigma}{2} & \text{for } \tau > t+h \end{cases}, \quad (34)$$

while $\theta_{i,\tau} = 0$ for all $i \geq 1$.

It is possible to show that the natural extensions of the inequality results established in Corollaries 1, 2, 3, 4 and 5 also hold for this class of processes:

Proposition 1 *Define the following variables*

$$\begin{aligned}
\overline{MSE}_T(h) &\equiv \frac{1}{T} \sum_{t=1}^T MSE_t(h), \text{ where } MSE_t(h) \equiv E \left[\left(Y_t - \hat{Y}_{t|h}^* \right)^2 \right] \\
\overline{MSF}_T(h) &\equiv \frac{1}{T} \sum_{t=1}^T MSF_t(h), \text{ where } MSF_t(h) \equiv E \left[\hat{Y}_{t|h}^{*2} \right], \\
\overline{C}_T(h) &\equiv \frac{1}{T} \sum_{t=1}^T C_t(h), \text{ where } C_t(h) \equiv E \left[\hat{Y}_{t|h}^* Y_t \right] \\
\overline{MSFR}_T(h_S, h_L) &\equiv \frac{1}{T} \sum_{t=1}^T MSFR_t(h), \text{ where } MSFR_t(h_S, h_L) \equiv E \left[\eta_{t|h_S, h_L}^2 \right] \\
\overline{B}_T(h) &\equiv \frac{1}{T} \sum_{t=1}^T B_t(h), \text{ where } B_t(h_S, h_L) \equiv E \left[Y_t \eta_{t|h_S, h_L} \right]
\end{aligned}$$

Under the conditions of Theorem 1 and S2, for any $h_S < h_M < h_L$ we then obtain the following :

$$\begin{aligned}
(a) \quad & \overline{MSE}_T(h_S) \leq \overline{MSE}_T(h_L) \\
(b) \quad & \overline{MSF}_T(h_S) \geq \overline{MSF}_T(h_L) \\
(c) \quad & \overline{C}_T(h_S) \geq \overline{C}_T(h_L) \\
(d) \quad & \overline{MSFR}_T(h_S, h_M) \leq \overline{MSFR}_T(h_S, h_L) \\
(e) \quad & \overline{MSFR}_T(h_S, h_L) \leq 2\overline{B}_T(h_S, h_L)
\end{aligned}$$

The inequalities for averages of unconditional moments presented in Proposition 1 can be tested by drawing on a central limit theorem for heterogeneous, serially dependent processes, see Wooldridge and White (1988) and White (2001) for example. The following proposition provides conditions under which these quantities can be estimated.

Proposition 2 *Define*

$$\begin{aligned}
d_{ht} &\equiv \left(Y_t - \hat{Y}_{t|h}^* \right)^2 - \left(Y_t - \hat{Y}_{t|t-(h-1)}^* \right)^2, \text{ for } h = h_2, \dots, h_H \\
\mathbf{d}_t &\equiv [d_{2t}, \dots, d_{Ht}]', \quad \hat{\Delta}_T \equiv \frac{1}{T} \sum_{t=1}^T \mathbf{d}_t', \quad V_T \equiv V \left[\frac{1}{\sqrt{T}} \sum_{t=1}^T \mathbf{d}_t' \right]
\end{aligned}$$

Then assume: (i) $\mathbf{d}_t = \Delta + \epsilon_t$, for $t = 1, 2, \dots$, $\Delta \in \mathbb{R}^{H-1}$; (ii) ϵ_t is a mixing sequence with either ϕ of size $-r/2(r-1)$, $r \geq 2$ or α of size $-r/(r-2)$, $r > 2$; (iii) $E[\epsilon_t] = 0$ for $t = 1, 2, \dots$; (iv) $E[|\epsilon_{it}|^r] < C < \infty$ for $i = 1, 2, \dots, H-1$; (v) V_T is uniformly positive definite; (vi) There exists a \hat{V}_T that is symmetric and positive definite such that $\hat{V}_T - V_T \rightarrow^p 0$. Then:

$$\hat{V}_T^{-1/2} \sqrt{T} \left(\hat{\Delta}_T - \Delta \right) \Rightarrow N(0, I) \text{ as } T \rightarrow \infty.$$

Thus we can estimate the average of unconditional moments with the usual sample average, with the estimator of the covariance matrix suitably adjusted, and then conduct the test of inequalities using Wolak's (1989) approach.

4.3 Model misspecification

A natural question to ask is whether the forecast optimality tests presented so far are really tests of forecast optimality or, rather, test that forecasters use consistent models in updating their forecasts as the horizon changes. As we indicated earlier, the optimality tests that rely exclusively on the forecasts—and thus exclude information on the outcome variable—test for consistency in forecasters' revisions. This raises the question whether our tests are valid if forecasters use different and possibly misspecified models at different horizons—a situation that might arise if forecasters use the 'direct' approach of fitting separate models to each forecast horizon as opposed to the 'iterated' approach where a single model is fitted to the shortest horizon and then iterated on to obtain multi-step forecasts.

The variance bounds remain valid in situations where forecasters use misspecified models. They require, however, that forecasters realize if they are using a suboptimal short-horizon model whose predictions are dominated by the forecasts from a long-horizon model. Consider, for example, the hypothetical situation where the (misspecified) one-step forecasting model delivers less precise forecasts than, say, a two-step forecasting model. The variance bound results then require that forecasters realize this and either switch to using the two-period forecasts outright (as in the example below) or improve upon their one-step forecasting model.

To illustrate this, consider a simple MA(2) specification:⁴

$$Y_t = \varepsilon_t + \theta \varepsilon_{t-2}, \quad (35)$$

where $\varepsilon_t \sim (0, \sigma_\varepsilon^2)$ is white noise, and suppose that forecasters generate h -step forecasts by regressing Y_t on Y_{t-h} . In particular, one-period forecasts are based on the model

$$Y_t = \beta_1 y_{t-1} + u_t.$$

It is easily verified for this process that $p \lim(\hat{\beta}_1) = 0$, so $\hat{Y}_{t|t-1} = 0$, and $E[(Y_t - \hat{Y}_{t|t-1})^2] = \sigma_\varepsilon^2(1 + \theta^2)$.

⁴We are grateful to Ken West for proposing this example.

Turning next to the two-period forecast horizon, suppose the forecaster regresses Y_t on Y_{t-2} ,

$$Y_t = \beta_2 Y_{t-2} + u_t,$$

where $p \lim(\hat{\beta}_2) = \theta/(1 + \theta^2)$. This means that $\hat{Y}_{t|t-2} = \theta(\varepsilon_{t-2} + \theta\varepsilon_{t-4})/(1 + \theta^2)$, so

$$\begin{aligned} E[(Y_t - \hat{Y}_{t|t-2})^2] &= E \left[\left(\varepsilon_t + \frac{\theta(1 + \theta^2) - \theta}{1 + \theta^2} \varepsilon_{t-2} - \frac{\theta^2}{1 + \theta^2} \varepsilon_{t-4} \right)^2 \right] \\ &= \sigma_\varepsilon^2 \left(1 + \frac{\theta^6 + \theta^4}{(1 + \theta^2)^2} \right) = \sigma_\varepsilon^2 \left(1 + \frac{\theta^4}{1 + \theta^2} \right). \end{aligned}$$

It is easily seen that, in this case,

$$E[(Y_t - \hat{Y}_{t|t-2})^2] \leq E[(Y_t - \hat{Y}_{t|t-1})^2],$$

seemingly in contradiction of the MSE inequality (2). The reason the MSE inequality fails in this example is that optimizing forecasters should realize, at time $t-1$, that they are using a misspecified model and that in fact the forecast from the previous period, $\hat{Y}_{t|t-2}$, produces a lower MSE. Hence, in this example a better forecast at time $t-1$ is $\hat{Y}_{t|t-1} = \hat{Y}_{t|t-2} = \theta(\varepsilon_{t-2} + \theta\varepsilon_{t-4})/(1 + \theta^2)$.

Notice also that the mean squared forecast variance bound (6) is violated here since

$$\text{var}(\hat{Y}_{t|t-1}) = 0 < \text{var}(\hat{Y}_{t|t-2}) = \sigma_\varepsilon^2 \theta^2 / (1 + \theta^2).$$

Thus, clearly the forecaster in this example is not producing optimal forecasts.

This example also illustrates that our variance bounds can be used to identify suboptimal forecasts and hence help to improve on misspecified models.

In some special situations, by virtue of being weak inequalities, the variance inequality tests may not have power to detect a sub-optimal forecast. This situation arises, for example, when all forecasts are constant, i.e. $\hat{Y}_{t|t-h} = c$ for all h . For this “broken clock” forecast, MSE-values are constant across horizons, forecast variances are zero as are the forecast revisions and the covariance between forecasts and actual values. This is a very special case, however, that rules out any variation in the forecast and so can be deemed empirically irrelevant.

5 Monte Carlo Simulations

There is little existing evidence on the finite sample performance of forecast optimality tests, particularly when multiple forecast horizons are simultaneously involved. Moreover, we have proposed

a set of new optimality tests which take the form of bounds on second moments of the data and require using the Wolak (1989) test of inequality constraints which also has not been widely used so far.⁵ For these reasons it is important to shed light on the finite sample performance of the various forecast optimality tests. Unfortunately, obtaining analytical results on the size and power of these tests for realistic sample sizes and types of alternatives is not possible. To overcome this, we use Monte Carlo simulations of a variety of different scenarios. We next describe the simulation design and then present the size and power results.

5.1 Simulation design

To capture persistence in the underlying data, we consider a simple AR(1) model for the data generating process:

$$\begin{aligned} Y_t &= \mu_y + \phi (Y_{t-1} - \mu_y) + \varepsilon_t, \quad t = 1, 2, \dots, T = 100 \\ \varepsilon_t &\sim iid N(0, \sigma_\varepsilon^2). \end{aligned} \tag{36}$$

We calibrate the parameters to quarterly US CPI inflation data:

$$\phi = 0.5, \quad \sigma_y^2 = 0.5, \quad \mu_y = 0.75.$$

Optimal forecasts for this process are given by:

$$\begin{aligned} \hat{Y}_{t|t-h}^* &= E_{t-h}[Y_t] \\ &= \mu_y + \phi^h (Y_{t-h} - \mu_y) \end{aligned}$$

We consider all horizons between $h = 1$ and H , and we set $H \in \{4, 8\}$.

5.1.1 Measurement error

The performance of optimality tests that rely on the target variable versus tests that only use forecasts is likely to be heavily influenced by measurement errors in the underlying target variable, Y_t . To study the effect of this, we assume that the target variable, \tilde{Y}_t , is observed with error, ψ_t

$$\tilde{Y}_t = Y_t + \psi_t, \quad \psi_t \sim iid N(0, \sigma_\psi^2).$$

⁵One exception is Patton and Timmermann (2009) who provide some evidence on the performance of the Wolak test in the context of tests of financial return models.

We consider three values for the magnitude of the measurement error, σ_ψ , calibrated relative to the standard deviation of the underlying variable, σ_y , namely (i) *zero*, $\sigma_\psi = 0$ (as for CPI); (ii) *medium*, $\sigma_\psi/\sigma_y = 0.65$ (as for GDP growth first release data);⁶ and (iii) *high*, $\sigma_\psi/\sigma_y = 1$.

5.1.2 Sub-optimal forecasts

To study the power of the optimality tests, we consider a variety of ways in which the forecasts can be suboptimal. First, we consider forecasts that are contaminated by the same level of noise at all horizons:

$$\hat{Y}_{t|t-h} = \hat{Y}_{t|t-h}^* + \sigma_{\xi,h} Z_{t,t-h}, \quad Z_{t,t-h} \sim iid N(0,1),$$

where $\sigma_{\xi,h} = 0.65\sigma_y$ for all h and thus has the same magnitude as the medium level measurement error.

Forecasts may alternatively be affected by noise whose standard deviation is *increasing* in the horizon, ranging from zero for the short-horizon forecast to $2 \times 0.65\sigma_y$ for the longest forecast horizon ($H = 8$):

$$\sigma_{\xi,h} = \frac{2(h-1)}{7} \times 0.65\sigma_y, \quad \text{for } h = 1, 2, \dots, H \leq 8.$$

Forecasts affected by noise whose magnitude is *decreasing* in the horizon (from zero for $h = 8$ to $2 \times 0.65\sigma_y$ for $h = 1$) take the form:

$$\sigma_{\xi,h} = \frac{2(8-h)}{7} \times 0.65\sigma_y, \quad \text{for } h = 1, 2, \dots, H \leq 8.$$

Finally, consistent with example 3 we consider forecasts with either “sticky” updating or, conversely, “overshooting”:

$$\hat{Y}_{t|t-h} = \gamma \hat{Y}_{t|t-h}^* + (1-\gamma) \hat{Y}_{t|t-h-1}^*, \quad \text{for } h = 1, 2, \dots, H.$$

To capture “sticky” forecasts we set $\gamma = 1/2$, whereas for “overshooting” forecasts we set $\gamma = 1.5$.

Tests based on forecast revisions may have better finite-sample properties than tests based on the forecasts themselves, particularly when the underlying process is highly persistent.

⁶The “medium” value is calibrated to match US GDP growth data, as reported by Faust, Rogers and Wright (2005).

5.2 Results from the simulation study

Table 1 reports the size of the various tests for a nominal size of 10%. Results are based on 1,000 Monte Carlo simulations and a sample of 100 observations. The variance bounds tests are clearly under-sized, particularly for $H = 4$, where none of the tests have a size above 4%. In contrast, the MZ Bonferroni bound is over-sized.⁷ The vector MZ test is also hugely oversized, while the size of the univariate optimal revision regression is close to the nominal value of 10%. Because of the clear size distortions to the MZ Bonferroni bound and the vector MZ regression, we do not further consider those tests in the simulation study.

Turning to the power of the various forecast optimality tests, Table 2 reports the results of our simulations, using the three measurement noise scenarios (constant, increasing and decreasing noise) and the sticky updating and overshooting schemes, respectively. In the first scenario with equal noise across different horizons (Panel A), neither the MSE, MSF, MSFR or decreasing covariance bounds have much power to detect deviations from forecast optimality. This holds across all three levels of measurement error. In contrast, the covariance bound on forecast revisions has excellent power to detect this type of deviation from optimality—close to 100%—particularly when the short-horizon forecast, $\hat{Y}_{t|t-1}$, which is not affected by noise, is used as the dependent variable.⁸ The power is somewhat weaker when the covariance bound test is adopted on the actual variable, although it improves by roughly 10% when the measurement error is reduced from the high value to zero. The univariate optimal revision regression (21) also has excellent power properties, particularly when the dependent variable is the short-horizon forecast.

The scenario with additive measurement noise that increases in the horizon, h , is ideal for the decreasing MSF test since now the variance of the long-horizon forecast is artificially inflated in contradiction of (6). Thus, as expected, Panel B of Table 2 shows that this test has very good power under this scenario: 42% in the case with four forecast horizons, rising to 100% in the case with eight forecast horizons. The MSE and MSFR bounds have zero power for this type of deviation from forecast optimality. The covariance bound based on the predicted variable has power around

⁷Conventionally, Bonferroni bounds tests are conservative and tend to be undersized. Here, the individual MZ regression tests are even more oversized than shown here, so the Bonferroni bound leads to a reduction in the size of the test which is still oversized.

⁸The covariance bound (14) works so well because noise in the forecast increases $E[\eta_{t|h_S, h_L}^2]$ without affecting $E[Y_t \eta_{t|h_S, h_L}]$, thereby making it less likely that $E[Y_t \eta_{t|h_S, h_L} - \eta_{t|h_S, h_L}^2] \geq 0$ holds.

15% when $H = 4$, which increases to a power of 90-95% when $H = 8$. The covariance bound with the actual value replaced by the short-run forecast, (16), has the highest power among all tests, with power of 70% when $H = 4$ and power of 100% when $H = 8$. This is substantially higher than the power of the univariate optimal revision regression test (21) which has power close to 10-15% when conducted on the actual values and power of 60% when the short-run forecast is used as the dependent variable.⁹

Panel C of Table 2 shows that the scenario with noise in the forecast that decreases as a function of the forecast horizon, h , gives rise to high power for the increasing MSE test and also for the MSFR test with power again being much higher when $H = 8$ compared to when $H = 4$. However, once again the covariance bound test and the univariate optimal revision regression (21) have superior power.

The univariate optimal revision regression (21) and the covariance bound test have stronger power in the scenarios with noise that is either constant or decreasing in the forecast horizon, h , because the precision of the forecasts is much better at short horizons. Hence, the greater the noise that is added to the short horizon forecasts, the better these tests are able to detect inefficiency of the forecast.

The next scenario assumes sticky updating in the forecasts. In this case, shown in Panel D of Table 2, only the univariate optimal revision regression (21) seems to have much power to detect deviations from a fully rational forecast. The power lies between 27% and 60% for this test when the regression is based on the actual variable, and grows stronger as a result of reducing the measurement error from the high value to zero. Interestingly, power is close to 100% when the univariate optimal revision regression is based on the short-term forecasts.

In the final scenario with overshooting, shown in Panel E of Table 2, none of the tests has particularly high power. However, the univariate optimal revision regression dominates with power ranging between 23% and 58%, depending on the level of the measurement error and on how many horizons are included. In this case power is actually stronger, the fewer constraints (H) are considered.

We also consider using a Bonferroni bound to combine various tests based on actual values, forecasts only or all tests. Results for these tests are shown at the bottom of tables 1 and 2.

⁹For this case, $\hat{Y}_{t|t-h_H}$ is very poor, but this forecast is also very noisy and so deviations from rationality can be relatively difficult to detect.

In all cases we find that the size of the tests falls well below the nominal size, as expected for a Bonferroni-based test, although the power seems to be quite high and comparable to the best among the individual tests.

In conclusion, viewed across all four scenarios, the covariance bound test performs best among all the second-moment bounds. Interestingly, it generally performs much better than the MSE bound which is the most commonly known variance bound. Among the regression tests, excellent performance is found for the univariate optimal revision regression, particularly when the test uses the short-run forecast as the dependent variable. This test tends to have superior power properties and performs well across most deviations from forecast efficiency. Either the covariance test or the univariate optimal revision regression have the highest power in all the experiments considered here with the covariance bound test being best in the realistic case where the noise in the forecasts increases with the horizon. Bonferroni bound tests conducted on the regression and second moment tests are also found to have good properties, pooling the information across various individual tests.

6 Empirical Application

As an empirical illustration of the forecast optimality tests, we next evaluate the Federal Reserve “Greenbook” forecasts of GDP growth, the GDP deflator and CPI inflation. Data are from Faust and Wright (2009), who carefully extracted the Greenbook forecasts and actual values from real-time Fed publications.¹⁰ We use quarterly observations of the target variable over the period from 1982Q1 to 2000Q4. Forecasts begin with the current quarter and run up to eight quarters ahead in time. However, since the forecasts have many missing observations at the longest horizons and we are interested in aligning the data in “event time”, we only study horizons up to five quarters, i.e., $h = 0, 1, 2, 3, 4, 5$. A few quarterly observations are missing, leaving a total of 69 observations.

Empirical results are reported in Table 3. The key findings are as follows. For GDP growth we observe a strong rejection of internal consistency via the univariate optimal revision regression that uses the short-run forecast as the target variable, (24), and a mild violation of the increasing mean-squared forecast revision test, (11).

Turning to the GDP deflator, we find that several tests reject forecast optimality. In particular, the tests for a decreasing covariance, the covariance bound on forecast revisions, a decreasing

¹⁰We are grateful to Jonathan Wright for providing the data.

mean squared forecast, and the univariate optimal revision regression test all lead to rejections. Figure 1 illustrates the rejection of the variance bound based on the forecasts and shows that, in contradiction with (6) the MSF is not weakly decreasing in the horizon, h . In fact, the MSF is higher for $h = 5$ than for $h = 0$.

Finally, for the CPI inflation rate we find a violation of the bound on the variance of the revisions, (11), and a rejection through the univariate optimal revision regression.

For all three variables, the Bonferroni-based combination test rejects multi-horizon forecast optimality at the 5% level. The type of rejections gives some clues as to possible sources of sub-optimality.

The source of some of the rejections of forecast optimality is further illustrated in Figures 1-3. For each of the series, Figure 1 plots the mean squared errors and variance of the forecasts on top of each other. Under the null of forecast optimality, the forecast and forecast error should be orthogonal and the sum of these two components should be constant across horizons. Clearly, this does not hold here, particularly for the GDP deflator and CPI inflation series. As shown in Figure 2—which plots the mean squared error and forecast variances separately—the variance of the forecast in fact increases in the horizon for the GDP deflator, and it follows an inverse U -shaped pattern for CPI inflation, both in apparent contradiction of the decreasing forecast variance property established earlier.

Figure 3 plots mean squared errors and mean squared forecast revisions against the forecast horizon. Whereas the mean squared forecast revisions are mostly increasing as a function of the forecast horizon for the two inflation series, for GDP growth we observe the opposite pattern, namely a very high mean squared forecast revision at the one-quarter horizon, followed by lower values at longer horizons. This is the opposite of what we would expect and so explains the (weak) rejection of forecast optimality for this case.

The Monte Carlo simulations are closely in line with our empirical findings. Rejections of forecast optimality come mostly from the covariance bound (14), (16) and the univariate optimal revision regressions (21), (24). Moreover, for GDP growth, rejections tend to be stronger when only the forecasts are used. This makes sense since this variable is likely to be most affected by data revisions and measurement errors.

7 Conclusion

In this paper we propose several new tests of forecast optimality that exploit information from multi-horizon forecasts. Our new tests are based on (weak) monotonicity properties of second moment bounds that must hold across forecast horizons and so are joint tests of optimality across several horizons. We show that monotonicity tests, whether conducted on the squared forecast errors, squared forecasts, squared forecast revisions or the covariance between the target variable and the forecast revision can be restated as inequality constraints on regression models and that econometric methods proposed by Gouriéroux et al. (1982) and Wolak (1987, 1989) can be adopted. Suitably modified versions of these tests conducted on the sequence of forecasts or forecast revisions recorded at different horizons can be used to test the internal consistency properties of an optimal forecast, thereby side-stepping the issues that arise for conventional tests when the target variable is either missing or observed with measurement error.

Simulations suggest that the new tests are more powerful than extant ones and also have better finite sample size. In particular a new covariance bound test that constrains the variance of forecast revisions by their covariance with the outcome variable and a univariate joint regression test that includes the long-horizon forecast and all interim forecast revisions generally have good power to detect deviations from forecast optimality. These results show the importance of testing the joint implications of forecast rationality across multiple horizons when data is available. An empirical analysis of the Fed’s Greenbook forecasts of inflation and output growth corroborates the ability of the new tests to detect evidence of deviations from forecast optimality.

Our analysis in this paper assumed squared error loss. However, many of the results can be extended to allow for more general loss functions with known shape parameters. For example, the MSE bound is readily generalized to a bound based on non-decreasing expected loss as the horizon grows, see Patton and Timmermann (2007a). Similarly, the orthogonality regressions can be extended to use the generalized forecast error, which is essentially the score associated with the forecaster’s first order condition, see Patton and Timmermann (2010). Allowing for the case with a parametric loss function but unknown (estimated) parameters is more involved and is a topic we leave for future research.

8 Appendix: Proofs

Proof of Corollary 1. By the optimality of $\hat{Y}_{t|t-h_S}^*$ and that $\hat{Y}_{t|t-h_L}^* \in \mathcal{F}_{t-h_S}$ we have $E_{t-h_S} \left[\left(Y_t - \hat{Y}_{t|t-h_S}^* \right)^2 \right] \leq E_{t-h} \left[\left(Y_t - \hat{Y}_{t|t-h_L}^* \right)^2 \right]$, which implies $E \left[\left(Y_t - \hat{Y}_{t|t-h_S}^* \right)^2 \right] \leq E \left[\left(Y_t - \hat{Y}_{t|t-h_L}^* \right)^2 \right]$ by the LIE, and so $MSE(h_S) \leq MSE(h_L)$. ■

Proof of Corollary 2. Forecast optimality under MSE loss implies $\hat{Y}_{t|t-h}^* = E_{t-h} [Y_t]$. Thus $E_{t-h} [e_{t|t-h}^*] \equiv E_{t-h} [Y_t - \hat{Y}_{t|t-h}^*] = 0$, which implies $E [e_{t|t-h}^*] = 0$ and $Cov [\hat{Y}_{t|t-h}^*, e_{t|t-h}^*] = 0$, and so $V [Y_t] = V [\hat{Y}_{t|t-h}^*] + E [e_{t|t-h}^{*2}]$, or $V [\hat{Y}_{t|t-h}^*] = V [Y_t] - E [e_{t|t-h}^{*2}]$. Corollary 1 showed that $E [e_{t|t-h}^{*2}]$ is weakly increasing in h , which implies that $V [\hat{Y}_{t|t-h}^*]$ must be weakly decreasing in h . Finally, note that $V [\hat{Y}_{t|t-h}^*] = E [\hat{Y}_{t|t-h}^{*2}] - E [\hat{Y}_{t|t-h}^*]^2 = E [\hat{Y}_{t|t-h}^{*2}] - E [Y_t]^2$, since $E [\hat{Y}_{t|t-h}^*] = E [Y_t]$. Thus if $V [\hat{Y}_{t|t-h}^*]$ is weakly decreasing in h we also have that $E [\hat{Y}_{t|t-h}^{*2}]$ is weakly decreasing in h . ■

Proof of Corollary 3. As used in the above proofs, forecast optimality implies $Cov [\hat{Y}_{t|t-h}^*, e_{t|t-h}^*] = 0$ and thus $Cov [\hat{Y}_{t|t-h}^*, Y_t] = Cov [\hat{Y}_{t|t-h}^*, \hat{Y}_{t|t-h}^* + e_{t|t-h}^*] = V [\hat{Y}_{t|t-h}^*]$. Corollary 2 showed that $V [\hat{Y}_{t|t-h}^*]$ is weakly decreasing in h , and thus we have that $Cov [\hat{Y}_{t|t-h}^*, Y_t]$ is also weakly decreasing in h . Further, since $Cov [\hat{Y}_{t|t-h_S}^*, Y_t] = E [\hat{Y}_{t|t-h_S}^* Y_t] - E [\hat{Y}_{t|t-h_S}^*] E [Y_t] = E [\hat{Y}_{t|t-h_S}^* Y_t] - E [Y_t]^2$, we also have that $E [\hat{Y}_{t|t-h_S}^* Y_t]$ is weakly decreasing in h . ■

Proof of Corollary 4. $\eta_{t|h_S, h_L} \equiv \hat{Y}_{t|t-h_S}^* - \hat{Y}_{t|t-h_L}^* = \eta_{t|h_S, h_L} \equiv (\hat{Y}_{t|t-h_S}^* - \hat{Y}_{t|t-h_M}^*) + (\hat{Y}_{t|t-h_M}^* - \hat{Y}_{t|t-h_L}^*) \equiv \eta_{t|h_S, h_M} + \eta_{t|h_M, h_L}$. Under the assumption that $h_S < h_M < h_L$ note that $E_{t-h_M} [\eta_{t|h_S, h_M}] = E_{t-h_M} [\hat{Y}_{t|t-h_S}^* - \hat{Y}_{t|t-h_M}^*] = 0$ by the law of iterated expectations. Thus $Cov [\eta_{t|h_S, h_M}, \eta_{t|h_M, h_L}] = 0$ and so $V [\eta_{t|h_S, h_L}] = V [\eta_{t|h_S, h_M}] + V [\eta_{t|h_M, h_L}] \geq V [\eta_{t|h_S, h_M}]$. Further, since $E_{t-h} [\eta_{t|h, k}] = 0$ for any $h < k$ we then have $E [\eta_{t|h, k}] = 0$ and thus $E [\eta_{t|h_S, h_L}^2] \geq E [\eta_{t|h_S, h_M}^2]$. ■

Proof of Corollary 5. For any $h_S < h_L$, Corollary 1 showed

$$\begin{aligned}
V \left[Y_t - \hat{Y}_{t|h_L}^* \right] &\geq V \left[Y_t - \hat{Y}_{t|h_S}^* \right] \\
\text{so } V \left[Y_t \right] + V \left[\hat{Y}_{t|h_L}^* \right] - 2Cov \left[Y_t, \hat{Y}_{t|h_L}^* \right] &\geq V \left[Y_t \right] + V \left[\hat{Y}_{t|h_S}^* \right] - 2Cov \left[Y_t, \hat{Y}_{t|h_S}^* \right] \\
\text{and } V \left[\hat{Y}_{t|h_L}^* \right] - 2Cov \left[Y_t, \hat{Y}_{t|h_L}^* \right] &\geq V \left[\hat{Y}_{t|h_S}^* \right] - 2Cov \left[Y_t, \hat{Y}_{t|h_S}^* \right] \\
&= V \left[\hat{Y}_{t|h_L}^* + \eta_{t|h_S, h_L} \right] - 2Cov \left[Y_t, \hat{Y}_{t|h_L}^* + \eta_{t|h_S, h_L} \right] \\
&= V \left[\hat{Y}_{t|h_L}^* \right] + V \left[\eta_{t|h_S, h_L} \right] \\
&\quad - 2Cov \left[Y_t, \hat{Y}_{t|h_L}^* \right] - 2Cov \left[Y_t, \eta_{t|h_S, h_L} \right].
\end{aligned}$$

Thus $V \left[\eta_{t|h_S, h_L} \right] \leq 2Cov \left[Y_t, \eta_{t|h_S, h_L} \right]$. ■

Proof of Corollary 6. (a) Let $h < k$, then $Cov \left[\hat{Y}_{t|k}^*, \hat{Y}_{t|h}^* \right] = Cov \left[\hat{Y}_{t|k}^*, \hat{Y}_{t|h}^* + \eta_{t|h, k} \right] = V \left[\hat{Y}_{t|h}^* \right]$, since $Cov \left[\hat{Y}_{t|k}^*, \eta_{t|h, k} \right] = 0$. From Corollary 2 we have that $V \left[\hat{Y}_{t|k}^* \right]$ is decreasing in k and thus $Cov \left[\hat{Y}_{t|k}^*, \hat{Y}_{t|h}^* \right]$ is decreasing in k . Since $E \left[\hat{Y}_{t|k}^* \right] = EV \left[Y_t \right]$ for all k , thus also implies that $E \left[\hat{Y}_{t|k}^* \hat{Y}_{t|h}^* \right]$ is decreasing in k .

(b) From Corollary 5 we have $V \left[\eta_{t|h_M, h_L} \right] \leq 2Cov \left[Y_t, \eta_{t|h_M, h_L} \right] = 2Cov \left[\hat{Y}_{t|h_S}^* + e_{t|h_S}^*, \eta_{t|h_M, h_L} \right] = 2Cov \left[\hat{Y}_{t|h_S}^*, \eta_{t|h_M, h_L} \right]$ since $Cov \left[e_{t|h_S}^*, \eta_{t|h_M, h_L} \right] = 0$. Further, since $E \left[\eta_{t|h_M, h_L} \right] = 0$ this also implies that $E \left[\eta_{t|h_M, h_L}^2 \right] \leq 2E \left[\hat{Y}_{t|h_S}^* \eta_{t|h_M, h_L} \right]$. ■

Proof of Corollary 7. The population value of β_h is $Cov \left[\hat{Y}_{t|h}^*, Y_t \right] / V \left[\hat{Y}_{t|h}^* \right]$, which under optimality equals $\beta_h = Cov \left[\hat{Y}_{t|h}^*, Y_t \right] / V \left[\hat{Y}_{t|h}^* \right] = Cov \left[\hat{Y}_{t|h}^*, \hat{Y}_{t|h}^* + e_{t|h}^* \right] / V \left[\hat{Y}_{t|h}^* \right] = V \left[\hat{Y}_{t|h}^* \right] / V \left[\hat{Y}_{t|h}^* \right] = 1$. The population value of α_h under optimality equals $\alpha_h = E \left[Y_t \right] - \beta_h E \left[\hat{Y}_{t|h}^* \right] = E \left[Y_t \right] - E \left[\hat{Y}_{t|h}^* \right] = 0$ by the LIE since $\hat{Y}_{t|h}^* = E_{t|h} \left[Y_t \right]$. ■

Proof of Corollary 8. First, we re-write the regression in equation (21) as a function of the individual forecasts, using the fact that $\eta_{t|h_j, h_{j+1}} \equiv \hat{Y}_{t|h_j} - \hat{Y}_{t|h_{j+1}}$

$$\begin{aligned}
Y_t &= \alpha + \beta_H \hat{Y}_{t|h_H} + \beta_1 \left(\hat{Y}_{t|h_{H-1}} - \hat{Y}_{t|h_H} \right) + \beta_2 \left(\hat{Y}_{t|h_{H-2}} - \hat{Y}_{t|h_{H-1}} \right) + \dots \\
&\quad + \beta_{H-1} \left(\hat{Y}_{t|h_1} - \hat{Y}_{t|h_H} \right) + u_t \\
&= \alpha + \beta_1 \hat{Y}_{t|h_1} + (\beta_2 - \beta_1) \hat{Y}_{t|h_{H-1}} + \dots + (\beta_H - \beta_{H-1}) \hat{Y}_{t|h_H} \\
&\equiv \alpha + \lambda_1 \hat{Y}_{t|h_1} + \lambda_2 \hat{Y}_{t|h_{H-1}} + \dots + \lambda_H \hat{Y}_{t|h_H}
\end{aligned}$$

We next use the Frisch-Waugh-Lovell theorem (see Davidson and MacKinnon, 1993, for example) to show that $\lambda_1 = 1$ and $\alpha = \lambda_2 = \dots = \lambda_H = 0$. Under the null of forecast optimality we

have $\hat{Y}_{t|t-h_1}^* = E_{t-h_1}[Y_t]$. Consider a first-stage regression of Y_t on $\hat{Y}_{t|t-h_1}^*$ (without a constant). Following the steps in the proof of Corollary 7, this yields a coefficient of one, and regression residuals are $e_{t|t-h_S}^* \equiv Y_t - \hat{Y}_{t|t-h_S}^*$. Next consider a regression of $\hat{Y}_{t|t-h_j}^*$ on $\hat{Y}_{t|t-h_1}^*$ (again without a constant) for each $j = 2, 3, \dots, H$, and let the slope coefficients from these regressions be denoted γ_j . Finally consider a regression of the residuals from the first regression on the matrix of residuals from the latter $H - 1$ regressions, namely:

$$e_{t|t-h_1}^* = \delta_0 + \delta_2 \left(\hat{Y}_{t|t-h_2}^* - \gamma_2 \hat{Y}_{t|t-h_1}^* \right) + \dots + \delta_H \left(\hat{Y}_{t|t-h_H}^* - \gamma_H \hat{Y}_{t|t-h_1}^* \right) + \xi_t.$$

Since forecast optimality implies $E_{t-h_1} \left[e_{t|t-h_1}^* \right] = 0$, we then have $\delta_0 = \delta_2 = \dots = \delta_H = 0$. By the Frisch-Waugh-Lovell theorem we know that $\delta_0 = \alpha$, $\delta_2 = \lambda_2, \dots, \delta_H = \lambda_H$. Thus the population values of $(\alpha, \lambda_2, \dots, \lambda_H)$ are zero, and the population value of λ_1 is thus one. This implies that $\alpha = 0$, $\beta_1 = 1$, $\beta_2 = 1, \dots, \beta_H = 1$, as claimed. ■

Proof of Corollary 9. (a) Under optimality, $\tilde{\beta}_h = \text{Cov} \left[\hat{Y}_{t|t-h_1}^*, \hat{Y}_{t|t-h_j}^* \right] / V \left[\hat{Y}_{t|t-h_j}^* \right] = \text{Cov} \left[\hat{Y}_{t|t-h_j}^* + \eta_{t|h_1, h_j}, \hat{Y}_{t|t-h_j}^* \right] / V \left[\hat{Y}_{t|t-h_j}^* \right] = V \left[\hat{Y}_{t|t-h_j}^* \right] / V \left[\hat{Y}_{t|t-h_j}^* \right] = 1$, and $\tilde{\alpha} = E \left[\hat{Y}_{t|t-h_1}^* \right] - \tilde{\beta} E \left[\hat{Y}_{t|t-h_j}^* \right] = E[Y_t] - E[Y_t] = 0$.

(b) Follows using the same steps as the proof of part (b) of Corollary 8, noting that $\hat{Y}_{t|t-h_2}^* = E_{t-h_2}[Y_t] = E_{t-h_2} \left[\hat{Y}_{t|t-h_1}^* \right]$ by the LIE, and that $E_{t-h_2} \left[\hat{Y}_{t|t-h_1}^* - \hat{Y}_{t|t-h_2}^* \right] \equiv E_{t-h_2} \left[\eta_{t|h_1, h_2} \right] = 0$. ■

Proof of Proposition 1. Throughout the proof we will use the fact that

$$\hat{Y}_{t|t-h}^* = E_{t-h}[Y_t] = E_{t-h} \left[f(t, \theta) + \sum_{i=0}^{\infty} \theta_{it} \varepsilon_{t-i} \right] = f(t, \theta) + \sum_{i=h}^{\infty} \theta_{it} \varepsilon_{t-i}$$

$$\text{and } e_{t|t-h}^* \equiv Y_t - \hat{Y}_{t|t-h}^* = \sum_{i=0}^{h-1} \theta_{it} \varepsilon_{t-i}$$

(a) From above we have

$$E \left[e_{t|t-h}^{*2} \right] = V \left[e_{t|t-h}^* \right] = V \left[\sum_{i=0}^{h-1} \theta_{it} \varepsilon_{t-i} \right] = \sigma_\varepsilon^2 \sum_{i=0}^{h-1} \theta_{it}^2$$

$$\text{thus } E \left[e_{t|t-h_L}^{*2} \right] - E \left[e_{t|t-h_S}^{*2} \right] = \sigma_\varepsilon^2 \sum_{i=0}^{h_L-1} \theta_{it}^2 - \sigma_\varepsilon^2 \sum_{i=0}^{h_S-1} \theta_{it}^2 = \sigma_\varepsilon^2 \sum_{i=h_S}^{h_L-1} \theta_{it}^2 \geq 0$$

which implies that $MSE_t(h_S) \leq MSE_t(h_L)$ for all t . This further implies that

$$\overline{MSE}_T(h_S) \equiv \frac{1}{T} \sum_{t=1}^T MSE_t(h_S) \leq \frac{1}{T} \sum_{t=1}^T MSE_t(h_L) \equiv \overline{MSE}_T(h_L) \text{ for all } T.$$

The proofs of (b)-(e) all follow similar extensions as used in part (a) and are omitted in the interests of brevity. ■

Proof of Proposition 2. Follows from Exercise 5.21 of White (2001). ■

References

- [1] Capistran, Carlos, 2007, Optimality Tests for Multi-Horizon Forecasts. Manuscript no 2007-14, Banco de Mexico.
- [2] Clements, Michael P., 1997, Evaluating the Rationality of Fixed-Event Forecasts. *Journal of Forecasting* 16, 225-239.
- [3] Clements, Michael P., 2009, Internal consistency of survey respondents' forecasts: Evidence based on the Survey of Professional Forecasters. In *The Methodology and Practice of Econometrics. A Festschrift in Honour of David F. Hendry.* eds. Jennifer L. Castle and Neil Shephard. Oxford University Press. Chapter 8, 206 - 226.
- [4] Clements, Michael P., and David F. Hendry, 1998, *Forecasting Economic Time Series.* Cambridge. Cambridge University Press.
- [5] Corradi, Valentina, Andres Fernandez, and Norman R. Swanson, 2009, Information in the Revision Process of Real-Time Datasets, *Journal of Business Economics and Statistics* 27, 455-467.
- [6] Croushore, Dean, 2006, Forecasting with Real-Time Macroeconomic Data. Pages 961-982 in G. Elliott, C. Granger and A. Timmermann (eds.) *Handbook of Economic Forecasting*, North Holland: Amsterdam.
- [7] Croushore, Dean and Tom Stark, 2001, A Real-Time Data Set for Macroeconomists. *Journal of Econometrics* 105, 111-130.
- [8] Davies, A. and K. Lahiri, 1995, A New Framework for Analyzing Survey Forecasts using Three-Dimensional Panel Data, *Journal of Econometrics*, 68, 205-227.
- [9] Diebold, Francis X., 2001, *Elements of Forecasting.* 2nd edition. Ohio: South-Western.
- [10] 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, 603-610.
- [11] Faust, Jon, John Rogers and Jonathan Wright, 2005, News and Noise in G-7 GDP Announcements. *Journal of Money, Credit and Banking* 37, 403-419.
- [12] Faust, Jon and Jonathan Wright, 2009, Comparing Greenbook and Reduced Form Forecasts using a Large Realtime Dataset. *Journal of Business and Economic Statistics* 27, 468-479.

- [13] Giacomini, Raffaella and Halbert White, 2006, Tests of Conditional Predictive Ability. *Econometrica* 74, 6, 1545-1578.
- [14] Gourieroux, C., A. Holly and A. Monfort, 1982, Likelihood Ratio Test, Wald Test, and Kuhn-Tucker Test in Linear Models with Inequality Constraints on the Regression Parameters. *Econometrica* 50, 63-80.
- [15] 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, 499-526.
- [16] Mincer, Jacob, and Victor Zarnowitz, 1969, The Evaluation of Economic Forecasts, in J. Mincer (ed.) *Economic Forecasts and Expectations*, National Bureau of Economic Research, New York.
- [17] Moon, Roger, Frank Schorfheide, Eleonora Granziera and Mihye Lee, 2009, Inference for VARs identified with sign restrictions. Mimeo, University of Southern California and University of Pennsylvania.
- [18] Newey, Whitney K., and Kenneth D. West, 1987, A Simple, Positive Semidefinite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix, *Econometrica*, 55, 703-708.
- [19] Nordhaus, William D., 1987, Forecasting Efficiency: Concepts and Applications. *Review of Economics and Statistics* 69, 667-674.
- [20] Patton, Andrew J., and Allan Timmermann, 2007a, Properties of optimal forecasts under asymmetric loss and nonlinearity. *Journal of Econometrics*, 140(2), 884-918.
- [21] Patton, Andrew J., and Allan Timmermann, 2007b, Testing Forecast Optimality under Unknown Loss. *Journal of the American Statistical Association* 102, 1172-1184.
- [22] Patton, Andrew J., and Allan Timmermann, 2008, Predictability of output growth and inflation: A multi-horizon survey approach. Unpublished manuscript, Duke and UCSD.
- [23] Patton, Andrew J., and Allan Timmermann, 2009, Monotonicity in Asset Returns; New Tests with Applications to the Term Structure, the CAPM, and Portfolio Sorts. Forthcoming in *Journal of Financial Economics*.
- [24] Patton, Andrew J. and Allan Timmermann, 2010, Generalized Forecast Errors, A Change of Measure and Forecast Optimality. Forthcoming in T. Bollerslev, J. Russell, and M. Watson, (eds.), *Volatility and Time Series Econometrics: Essays in Honour of Robert F. Engle*, Oxford University Press.
- [25] Schmidt, Peter, 1974, The Asymptotic Distribution of Forecasts in the Dynamic Simulation of an Econometric Model. *Econometrica* 42, 303-309.
- [26] Timmermann, Allan, 1993, How Learning in Financial Markets Generates Excess Volatility and Predictability in Stock Prices. *Quarterly Journal of Economics* 108(4), 1135-1145.
- [27] West, Kenneth D., 1996, Asymptotic Inference about Predictive Ability. *Econometrica* 64, 1067-84.

- [28] West, Kenneth D., and Michael W. McCracken, 1998, Regression Based Tests of Predictive Ability. *International Economic Review* 39, 817-840.
- [29] White, Halbert, 2001, *Asymptotic Theory for Econometricians*, Second Edition, Academic Press, San Diego.
- [30] Wolak, Frank A., 1987, An Exact Test for Multiple Inequality and Equality Constraints in the Linear Regression Model. *Journal of the American Statistical Association* 82, 782-793.
- [31] Wolak, Frank A., 1989, Testing Inequality Constraints in Linear Econometric Models. *Journal of Econometrics* 31, 205-235.
- [32] Wooldridge, Jeffrey M., and Halbert White, 1988, Some Invariance Principles and Central Limit Theorems for Dependent Heterogeneous Processes, *Econometric Theory*, 4, 210-230.

**Table 1: Monte Carlo simulation of size of the inequality tests
and regression-based tests of forecast optimality**

| <i>Meas. error variance:</i> | H = 4 | | | H = 8 | | |
|---------------------------------|--------------|------------|-------------|--------------|------------|-------------|
| | <i>High</i> | <i>Med</i> | <i>Zero</i> | <i>High</i> | <i>Med</i> | <i>Zero</i> |
| Inc MSE | 1.9 | 1.7 | 1.1 | 7.8 | 6.4 | 8.3 |
| Dec COV | 1.1 | 1.1 | 0.8 | 8.4 | 7.3 | 7.2 |
| COV bound | 2.2 | 1.2 | 0.4 | 2.3 | 1.4 | 0.8 |
| Dec MSF | 2.1 | 2.1 | 2.1 | 5.3 | 5.3 | 5.3 |
| Inc MSFR | 0.4 | 0.4 | 0.4 | 5.5 | 5.5 | 5.5 |
| Dec COV, h=1 | 0.9 | 0.9 | 0.9 | 6.4 | 6.4 | 6.4 |
| COV bound, h=1 | 3.6 | 3.6 | 3.6 | 4.6 | 4.6 | 4.6 |
| Inc MSE & Dec MSF | 1.5 | 1.3 | 0.8 | 8.3 | 8.2 | 9.1 |
| Inc MSE & Inc MSFR | 1.1 | 0.8 | 0.6 | 7.2 | 6.7 | 6.5 |
| Univar MZ, Bonferroni | 13.8 | 15.0 | 17.8 | 19.5 | 19.4 | 20.3 |
| Univar MZ, Bonferroni, h=1 | 16.0 | 16.0 | 16.0 | 19.2 | 19.2 | 19.2 |
| Vector MZ | 39.8 | 38.0 | 31.2 | 63.0 | 62.0 | 58.8 |
| Vector MZ, h=1 | 25.2 | 25.2 | 25.2 | 52.4 | 52.4 | 52.4 |
| Univar opt. revision regr. | 11.3 | 11.5 | 11.0 | 12.4 | 11.8 | 11.0 |
| Univar opt. revision regr., h=1 | 12.0 | 12.0 | 12.0 | 11.3 | 11.3 | 11.3 |
| Bonf, using actuals | 3.9 | 4.2 | 3.6 | 7.4 | 7.6 | 8.0 |
| Bonf, using forecasts only | 3.0 | 3.0 | 3.0 | 6.6 | 6.6 | 6.6 |
| Bonf, all tests | 3.6 | 3.5 | 2.2 | 7.6 | 7.5 | 6.2 |

Notes: This table presents the outcome of 1,000 Monte Carlo simulations of the size of various forecast optimality tests. Data is generated by a first-order autoregressive process with parameters calibrated to quarterly US CPI inflation data, i.e. $\phi = 0.5$, $\sigma_y^2 = 0.5$ and $\mu_y = 0.75$. We consider three levels of error in the measured value of the target variable (high, median and zero). Optimal forecasts are generated under the assumption that this process (and its parameter values) are known to forecasters. The simulations assume a sample of 100 observations and a nominal size of 10%. The inequality tests are based on the Wolak (1989) test and use simulated critical values based on a mixture of chi-squared variables. Rows with ‘ $h = 1$ ’ refer to cases where the one-period forecast is used in place of the predicted variable.

Table 2: Monte Carlo simulation of power of the inequality tests and regression-based tests of forecast optimality

| | H = 4 | | | H = 8 | | |
|--|--------------|------------|-------------|--------------|------------|-------------|
| <i>Meas. error variance:</i> | <i>High</i> | <i>Med</i> | <i>Zero</i> | <i>High</i> | <i>Med</i> | <i>Zero</i> |
| <i>PANEL A: Equal noise across all forecast horizons</i> | | | | | | |
| Inc MSE | 7.5 | 6.8 | 6.8 | 13.4 | 12.4 | 12.6 |
| Dec COV | 7.3 | 6.4 | 6.1 | 13.0 | 13.5 | 12.2 |
| COV bound | 73.7 | 79.6 | 83.2 | 74.8 | 78.7 | 83.3 |
| Dec MSF | 5.8 | 5.8 | 5.8 | 15.0 | 15.0 | 15.0 |
| Inc MSFR | 9.9 | 9.9 | 9.9 | 14.8 | 14.8 | 14.8 |
| Dec COV, h=1 | 8.9 | 8.9 | 8.9 | 15.4 | 15.4 | 15.4 |
| COV bound, h=1 | 98.0 | 98.0 | 98.0 | 99.1 | 99.1 | 99.1 |
| Inc MSE & Dec MSF | 7.8 | 7.6 | 6.9 | 27.3 | 26.7 | 26.0 |
| Inc MSE & Inc MSFR | 8.2 | 7.3 | 7.0 | 23.4 | 23.3 | 23.0 |
| Univar opt. revision regr. | 91.9 | 98.1 | 99.6 | 85.3 | 95.9 | 99.0 |
| Univar opt. revision regr., h=1 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| Bonf, using actuals | 83.4 | 94.0 | 98.5 | 78.7 | 89.8 | 96.7 |
| Bonf, using forecasts only | 100.0 | 100.0 | 100.0 | 99.9 | 99.9 | 99.9 |
| Bonf, all tests | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |

Notes: This table presents the outcome of 1,000 Monte Carlo simulations of the size of various forecast optimality tests. Data is generated by a first-order autoregressive process with parameters calibrated to quarterly US CPI inflation data, i.e. $\phi = 0.5$, $\sigma_y^2 = 0.5$ and $\mu_y = 0.75$. We consider three levels of error in the measured value of the target variable (high, median and zero). Optimal forecasts are generated under the assumption that this process (and its parameter values) are known to forecasters. Power is then studied against sub-optimal forecasts obtained as follows: A: forecasts are contaminated by the same level of noise across all horizons; B: forecasts are contaminated by noise that increases in the horizon; C: forecasts are contaminated by noise that decreases in the horizon; D: Forecasts are updated in a sticky manner; E: forecasts overshoot their optimal values. The simulations assume a sample of 100 observations and a nominal size of 10%. Rows with ‘ $h = 1$ ’ refer to cases where the one-period forecast is used in place of the predicted variable.

Table 2: Monte Carlo simulation of power of the inequality tests and regression-based tests of forecast optimality

| | H = 4 | | | H = 8 | | |
|--|--------------|------------|-------------|--------------|------------|-------------|
| <i>Meas. error variance:</i> | <i>High</i> | <i>Med</i> | <i>Zero</i> | <i>High</i> | <i>Med</i> | <i>Zero</i> |
| <i>PANEL B: Noise increases with the horizon</i> | | | | | | |
| Inc MSE | 0.2 | 0.2 | 0.0 | 0.0 | 0.0 | 0.0 |
| Dec COV | 3.3 | 3.0 | 2.8 | 13.5 | 12.8 | 12.2 |
| COV bound | 12.9 | 14.5 | 14.9 | 90.7 | 93.3 | 95.2 |
| Dec MSF | 42.3 | 42.3 | 42.3 | 100.0 | 100.0 | 100.0 |
| Inc MSFR | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Dec COV, h=1 | 4.9 | 4.9 | 4.9 | 12.9 | 12.9 | 12.9 |
| COV bound, h=1 | 69.2 | 69.2 | 69.2 | 100.0 | 100.0 | 100.0 |
| Inc MSE & Dec MSF | 25.5 | 25.3 | 23.2 | 99.8 | 99.8 | 99.8 |
| Inc MSE & Inc MSFR | 0.2 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Univar opt. revision regr. | 11.7 | 12.3 | 11.9 | 13.1 | 13.6 | 12.9 |
| Univar opt. revision regr., h=1 | 63.6 | 63.6 | 63.6 | 54.6 | 54.6 | 54.6 |
| Bonf, using actuals | 7.9 | 9.2 | 9.2 | 80.1 | 84.1 | 86.6 |
| Bonf, using forecasts only | 63.0 | 63.0 | 63.0 | 100.0 | 100.0 | 100.0 |
| Bonf, all tests | 54.7 | 54.7 | 54.4 | 100.0 | 100.0 | 100.0 |
| <i>PANEL C: Noise decreases with the horizon</i> | | | | | | |
| Inc MSE | 71.3 | 79.8 | 87.1 | 100.0 | 100.0 | 100.0 |
| Dec COV | 6.7 | 5.8 | 6.1 | 13.6 | 12.5 | 13.1 |
| COV bound | 99.5 | 99.8 | 99.9 | 99.5 | 99.8 | 99.9 |
| Dec MSF | 0.4 | 0.4 | 0.4 | 0.1 | 0.1 | 0.1 |
| Inc MSFR | 55.9 | 55.9 | 55.9 | 100.0 | 100.0 | 100.0 |
| Dec COV, h=1 | 10.6 | 10.6 | 10.6 | 17.6 | 17.6 | 17.6 |
| COV bound, h=1 | 99.8 | 99.8 | 99.8 | 99.2 | 99.2 | 99.2 |
| Inc MSE & Dec MSF | 50.8 | 59.6 | 69.1 | 100.0 | 100.0 | 100.0 |
| Inc MSE & Inc MSFR | 79.6 | 83.1 | 88.8 | 100.0 | 100.0 | 100.0 |
| Univar opt. revision regr. | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| Univar opt. revision regr., h=1 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| Bonf, using actuals | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| Bonf, using forecasts only | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| Bonf, all tests | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |

Notes: See notes to Panel A of this table on previous page.

Table 2: Monte Carlo simulation of power of the inequality tests and regression-based tests of forecast optimality

| | H = 4 | | | H = 8 | | |
|---------------------------------|--------------|------------|-------------|--------------|------------|-------------|
| <i>Meas. error variance:</i> | <i>High</i> | <i>Med</i> | <i>Zero</i> | <i>High</i> | <i>Med</i> | <i>Zero</i> |
| <i>PANEL D: Sticky updating</i> | | | | | | |
| Inc MSE | 2.0 | 1.4 | 1.1 | 6.8 | 6.0 | 7.6 |
| Dec COV | 2.1 | 2.1 | 2.1 | 9.2 | 8.7 | 8.5 |
| COV bound | 0.9 | 0.7 | 0.4 | 1.2 | 0.8 | 0.5 |
| Dec MSF | 4.8 | 4.8 | 4.8 | 8.2 | 8.2 | 8.2 |
| Inc MSFR | 0.0 | 0.0 | 0.0 | 4.8 | 4.8 | 4.8 |
| Dec COV, h=1 | 2.4 | 2.4 | 2.4 | 7.4 | 7.4 | 7.4 |
| COV bound, h=1 | 2.7 | 2.7 | 2.7 | 4.3 | 4.3 | 4.3 |
| Inc MSE & Dec MSF | 2.6 | 2.4 | 2.1 | 11.0 | 11.0 | 10.7 |
| Inc MSE & Inc MSFR | 0.8 | 0.9 | 0.9 | 7.9 | 7.0 | 8.3 |
| Univar opt. revision regr. | 33.2 | 44.8 | 59.1 | 27.5 | 35.2 | 49.9 |
| Univar opt. revision regr., h=1 | 99.5 | 99.5 | 99.5 | 98.9 | 98.9 | 98.9 |
| Bonf, using actuals | 15.2 | 22.0 | 37.4 | 14.9 | 19.1 | 28.9 |
| Bonf, using forecasts only | 97.5 | 97.5 | 97.5 | 93.3 | 93.3 | 93.3 |
| Bonf, all tests | 96.3 | 96.2 | 96.2 | 89.4 | 89.4 | 89.4 |
| <i>PANEL E: Over-shooting</i> | | | | | | |
| Inc MSE | 2.9 | 2.4 | 2.2 | 8.8 | 8.0 | 7.3 |
| Dec COV | 1.1 | 0.5 | 0.7 | 5.6 | 5.7 | 5.6 |
| COV bound | 5.4 | 4.6 | 4.8 | 4.9 | 4.7 | 4.9 |
| Dec MSF | 1.0 | 1.0 | 1.0 | 3.6 | 3.6 | 3.6 |
| Inc MSFR | 2.5 | 2.5 | 2.5 | 8.1 | 8.1 | 8.1 |
| Dec COV, h=1 | 1.0 | 1.0 | 1.0 | 6.6 | 6.6 | 6.6 |
| COV bound, h=1 | 7.7 | 7.7 | 7.7 | 8.5 | 8.5 | 8.5 |
| Inc MSE & Dec MSF | 1.2 | 0.9 | 0.7 | 6.6 | 7.3 | 6.5 |
| Inc MSE & Inc MSFR | 2.0 | 2.1 | 1.9 | 10.1 | 9.5 | 7.1 |
| Univar opt. revision regr. | 29.8 | 41.8 | 57.9 | 23.8 | 32.4 | 48.0 |
| Univar opt. revision regr., h=1 | 32.3 | 32.3 | 32.3 | 27.9 | 27.9 | 27.9 |
| Bonf, using actuals | 13.5 | 23.3 | 38.0 | 8.7 | 14.8 | 28.2 |
| Bonf, using forecasts only | 0.0 | 0.0 | 0.0 | 13.7 | 13.7 | 13.7 |
| Bonf, all tests | 10.3 | 17.1 | 31.3 | 11.9 | 14.5 | 22.1 |

Notes: See notes to Panel A of this table on previous page.

Table 3: Forecast optimality tests for Greenbook forecasts

| <i>Series:</i> | <i>Growth</i> | <i>Deflator</i> | <i>Inflation</i> |
|---------------------------------|---------------|-----------------|------------------|
| Inc MSE | 0.599 | 0.964 | 0.644 |
| Dec COV | 0.898 | 0.058* | 0.991 |
| COV bound | 0.498 | 0.000* | 0.009* |
| Dec MSF | 0.898 | 0.026* | 0.725 |
| Inc MSFR | 0.084* | 0.936 | 0.624 |
| Dec COV, h=1 | 0.802 | 0.075* | 0.795 |
| COV bound, h=1 | 0.216 | 0.010* | 0.656 |
| Inc MSE & Dec MSF | 0.934 | 0.126 | 0.616 |
| Inc MSE & Inc MSFR | 0.250 | 0.992 | 0.749 |
| Univar opt. revision regr. | 0.709 | 0.000* | 0.001* |
| Univar opt. revision regr., h=1 | 0.000* | 0.009* | 0.022* |
| Bonf, using actuals | 1.000 | 0.000* | 0.004* |
| Bonf, using forecasts only | 0.000* | 0.047* | 0.108 |
| Bonf, all tests | 0.000* | 0.001* | 0.012* |

Note: This table presents p-values from inequality- and regression tests of forecast optimality applied to quarterly Greenbook forecasts of GDP growth, the GDP deflator and CPI Inflation. The sample covers the period 1982Q1-2000Q4. Six forecast horizons are considered, i.e., $h = 0, 1, 2, 3, 4, 5$ and the forecasts are aligned in event time. The inequality tests are based on the Wolak (1989) test and use simulated critical values based on a mixture of chi-squared variables. Rows with ' $h = 1$ ' refer to cases where the one-period forecast is used in place of the predicted variable.

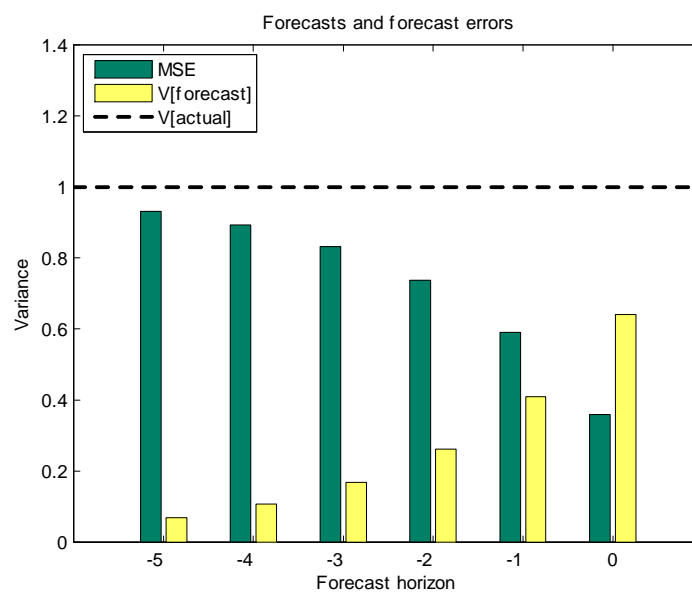


Figure 1: *Theoretical mean squared errors and forecast variances for an $AR(1)$ process with unconditional variance of 1 and autoregressive coefficient of 0.8.*

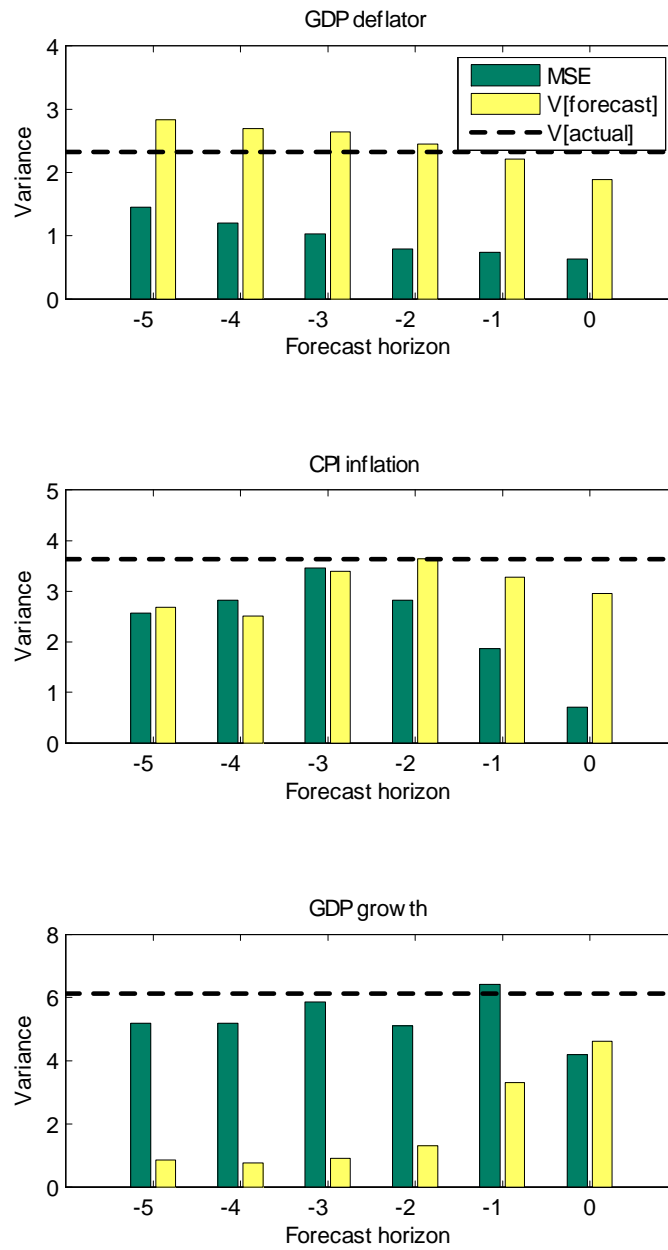


Figure 2: Mean squared errors and forecast variances, for US GDP deflator, CPI inflation and GDP growth.

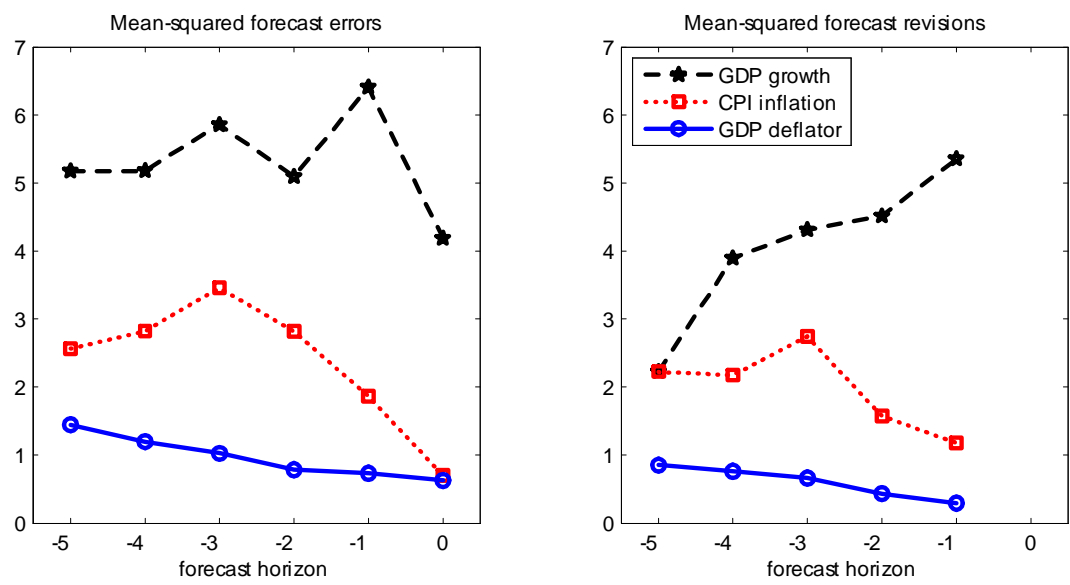


Figure 3: Mean squared errors (left panel) and mean-squared forecast revisions (right panel), for US GDP deflator, CPI inflation and GDP growth.