New Tests of Forecast Optimality Across Multiple Horizons

Andrew J. Patton — *Duke University*
Allan Timmermann — *University of California, San Diego*

17 February, 2010

We propose new tests of forecast optimality that exploit information contained in multi-horizon forecasts. In addition to implying zero forecast 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 at long and short 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 by Wolak (1989). Moreover, the tests can be conducted without the need for data on the target variable, which is particularly useful when this is subject to large measurement error. We also propose a new univariate test of forecast optimality that constrains the coefficients in a regression of the target variable on the long-horizon forecast and the sequence of interim forecast revisions. Size and power of the new tests are compared with those of conventional orthogonality 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

PRELIMINARY AND INCOMPLETE

1 Introduction

Forecasts recorded at multiple horizons, often from a single through several quarters ahead in time, 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 forecasts offer the potential of drawing more powerful conclusions about the ability of forecasters to produce optimal forecasts. This paper proposes a range of new forecast optimality tests and compares them with extant ones.

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 (2007). 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). Optimal updating of forecasts also 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 that are valid when forecasts across multiple horizons are simultaneously available. We propose new ways to construct such a test. Ordering squared forecast errors by forecast horizon, we show that the weak monotonicity property can be tested through a set of inequality constraints on a regression model using methods proposed by Gourieroux et al (1982) and Wolak (1987, 1989).

Tests of forecast optimality have conventionally been based on comparing predicted and "realized" 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 (2007), 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. Using only data on forecasts or forecast revisions, we show that both the new and extant optimality tests can be restated in a way that does not require having observations on the realized values. 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 forecasts gives rise to a generalized efficiency regression. Using a single equation, this test is undertaken by regressing either the realized value or the short-horizon forecast on an intercept, the long-horizon forecast and the sequence of intermediate forecast revisions. A set of zero-one restrictions on the intercept and slope coefficients are then tested. A key difference from the conventional efficiency test is that the joint consistency of all forecasts at different horizons gets 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 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 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 horizons 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 establishes the new variance bounds and sets up the associated hypothesis tests. 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 forecast optimality and/or internal consistency of a sequence of forecasts recorded at different horizons.

2.1 Assumptions and Setup

Consider a univariate time series, $Y \equiv \{Y_t; t = 1, 2, ...\}$, and suppose that forecasts of this variable are recorded at different points in time, $t = 1, ..., T$ and at different horizons, $h = h_1, ..., h_H$. 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 $\left\{\left(Y_{t-h-k}, \tilde{Z}'_{t-h-k}\right)'; k \geq 0\right\}$, where $\tilde{Z}_{t-h}$ is a vector of predictor variables capturing elements in the forecaster's information set at time $t-h$. 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 $h_1 < h_2 < \cdots < h_H$, with generic long and short horizons denoted by h_L and h_S $(h_L > h_S)$.

We assume that the forecaster has squared error loss, so the optimal forecast is obtained as:

$$\hat{Y}^*_{t|t-h} \equiv \underset{\hat{y} \in \mathcal{Y}}{\arg\min}\ E\left[(Y_t - \hat{y})^2 |\mathcal{F}_{t-h}\right]. \tag{1}$$

where $\mathcal{Y} \subseteq \mathbb{R}$ is the set of possible values for the forecast. From Granger (1969), we know that the optimal forecast is the conditional mean of the predicted variable, i.e.,

$$\hat{Y}^*_{t|t-h} = E[Y_t|\mathcal{F}_{t-h}]. \tag{2}$$

Moreover, the associated forecast errors, $e^*_{t|t-h} = Y_t - \hat{Y}^*_{t|t-h}$, should be mean-zero and uncorrelated with any $Z_{t-h} \in \mathcal{F}_{t-h}$. We next describe a variety of forecast optimality tests.

2.2 Monotonicity of mean squared errors

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

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

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:

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

By the law of iterated expectations and covariance stationarity of $Y_t, \hat{Y}^*_{t|t-h}$,

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

Assuming that forecasts are available at horizons $0 \leq h_1 \leq h_2 .... \leq h_H$, 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:[2]

$$E\left[e^{*2}_{t|t-h_1}\right] \leq E\left[e^{*2}_{t|t-h_2}\right] \leq ... \leq E\left[e^{*2}_{t|t-h_H}\right] \rightarrow V\left[Y_t\right] \text{ as } h_H \rightarrow \infty, \tag{3}$$

with inequalities being strict if more forecast-relevant information becomes available as the forecast horizon shrinks to zero.[3] This property is well-known as discussed by, e.g., Diebold (2001) and Patton and Timmermann (2007).

[1] We shall ignore the effect of parameter estimation error which could be important in finite samples. As shown by Schmidt (1974) and Clements and Hendry (1998), parameter estimation error introduces a term in the mean squared forecasts that is a non-monotonic function of the forecast horizon.

[2] Note that in the presence of measurement error and data revisions, the "forecast" horizon can be zero or negative (corresponding to "nowcasts" and "backcasts") and still generate positive MSE. Our results do not require the forecast horizon to be strictly positive.

[3] For 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.

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:

$$\tilde{Y}_{t|t-h_S} = \tilde{Y}_{t|t-h_L} + u_{t-h_S}, \quad u_{t-h_S} \perp \tilde{Y}_{t|t-h_L}, u_{t-h_S} \perp Y_t. \tag{4}$$

We then have

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

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, (3), 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 (3).

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

$$\Delta_h^e \equiv \mu_h - \mu_{h-1} = E\left[e_{t|t-h}^2\right] - E\left[e_{t|t-h+1}^2\right],$$

we can rewrite the inequalities in (3) as

$$\Delta_h^e \geq 0, \quad \text{for } h = 2, ..., H. \tag{5}$$

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:

$$\begin{aligned} H_0 &: \boldsymbol{\Delta}^e \geq \mathbf{0}, \\ \text{vs. } H_1 &: \boldsymbol{\Delta}^e \in \mathbb{R}^{H-1}, \end{aligned} \tag{6}$$

where the $(H-1) \times 1$ vector of MSE-differentials is given by $\boldsymbol{\Delta}^e \equiv [\Delta_2^e, ..., \Delta_H^e]'$. In contrast, $\boldsymbol{\Delta}^e$ is unconstrained under the alternative. Tests can be based on the sample analogs $\hat{\Delta}_h^e = \hat{\mu}_h - \hat{\mu}_{h-1}$ for $\hat{\mu}_h \equiv \frac{1}{T}\sum_{t=1}^{T} e_{t|t-h}^2$. 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*

When the target variable is not available, tests of monotonicity based on the mean squared forecast errors are of course infeasible. The derivations above do, however, lead to a test of optimality based only on the forecasts. Recalling that, under optimality, $E\left[e_{t|t-h}^*\right] = Cov\left[\hat{Y}_{t|t-h}^*, e_{t|t-h}^*\right] = 0$, we obtain:

$$\begin{aligned} V[Y_t] &= V\left[\hat{Y}_{t|t-h}^* + e_{t|t-h}^*\right] \\ &= V\left[\hat{Y}_{t|t-h}^*\right] + E\left[e_{t|t-h}^{*2}\right], \quad \text{so} \\ V\left[\hat{Y}_{t|t-h}^*\right] &= V[Y_t] - E\left[e_{t|t-h}^{*2}\right]. \end{aligned} \tag{7}$$

Thus 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.

We can transform this relation to one based on mean squared forecasts, which slightly simplifies the test (as it is then based on means rather than variances):

$$\begin{aligned} E\left[\hat{Y}_{t|t-h}^{*2}\right] &= V\left[\hat{Y}_{t|t-h}^*\right] + E\left[\hat{Y}_{t|t-h}^*\right]^2 \\ &= V\left[\hat{Y}_{t|t-h}^*\right] + E[Y_t]^2, \end{aligned}$$

since $E\left[\hat{Y}^*_{t|t-h}\right] = E[Y_t]$ under forecast optimality. Thus monotonicity of the variance of the forecast implies monotonicity of the mean-squared forecasts, i.e.,

$$E\left[\hat{Y}^{*2}_{t|t-h_1}\right] \geq E\left[\hat{Y}^{*2}_{t|t-h_2}\right] \geq \cdots \geq E\left[\hat{Y}^{*2}_{t|t-h_H}\right] \rightarrow E[Y_{t+h}]^2 \text{ as } h_H \rightarrow \infty. \tag{8}$$

A test of this implication can again be based on Wolak's (1989) approach by defining the vector $\boldsymbol{\Delta}^f \equiv \left[\Delta^f_2, ..., \Delta^f_H\right]'$, where $\Delta^f_h \equiv E\left[\hat{Y}^{*2}_{t|t-h}\right] - E\left[\hat{Y}^{*2}_{t|t-h+1}\right]$ and testing the null hypothesis that differences in mean squared forecasts are weakly negative for all forecast horizons:

$$\begin{aligned} H_0 &: \boldsymbol{\Delta}^f \leq \mathbf{0}, \\ \text{vs. } H_1 &: \boldsymbol{\Delta}^f \in \mathbb{R}^{H-1}. \end{aligned} \tag{9}$$

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.[4]

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:

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

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 $Cov\left[\hat{Y}^*_{t|t-h_L}, \eta_{t|h_S,h_L}\right] = 0$, and so:

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

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

[4] For other tests of internal consistency, see Clements (2009).

will have greater variance than the short-horizon forecast, and a test based on (8) 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 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 decreasing forecast variances as the forecast horizon expands is that the covariance of the forecasts with the target variable should be *decreasing* in the forecast horizon. To see this, note that

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

It follows that

$$E\left[\hat{Y}^*_{t|t-h} Y_t\right] = V\left[\hat{Y}^*_{t|t-h}\right] + E\left[Y_t\right]^2,$$

which should be decreasing in the forecast horizon since $V\left[\hat{Y}^*_{t|t-h}\right]$ is a declining function of h and $E\left[Y_t\right]^2$ is independent of the horizon.

As for the above cases, this can be tested using

$$\Delta^c_h \equiv E\left[\hat{Y}^*_{t|t-h} Y_t\right] - E\left[\hat{Y}^*_{t|t-h+1} Y_t\right], \text{ for } h = h_1, ..., h_H,$$

and considering the hypothesis

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

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:

$$\begin{aligned}\hat{Y}^*_{t|t-h_1} &= \hat{Y}^*_{t|t-h_H} + \left(\hat{Y}^*_{t|t-h_H+1} - \hat{Y}^*_{t|t-h_H}\right) + ... + \left(\hat{Y}^*_{t|t-h_1} - \hat{Y}^*_{t|t-h_2}\right) \\ &\equiv \hat{Y}^*_{t|t-h_H} + \sum_{j=1}^{H-1} \eta_{t|h_j,h_{j+1}}, \end{aligned} \tag{10}$$

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, $E_{t-h_L}\left[\eta_{t|h_S,h_L}\right] = 0$ for all $h_S < h_L$, so $Cov\left[\hat{Y}^*_{t|t-h_L}, \eta_{t|h_S,h_L}\right] = 0$, and

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

It follows that

$$V\left[\eta_{t|h_1,h_{H-1}}\right] \leq V\left[\eta_{t|h_1,h_H}\right], \tag{11}$$

or, equivalently,

$$E\left[\eta^2_{t|h_1,h_{H-1}}\right] \leq E\left[\eta^2_{t|h_1,h_H}\right].$$

Thus the variance of the forecast revision from time $t - h_{H-1}$ to time $t - h_1$ should be weakly less than that from time $t - h_H$ to $t - h_1$, for $h_{H-1} < h_H$. More generally, the mean squared forecast revision should be weakly increasing in the difference between the forecast horizons:

$$E\left[\eta^2_{t|h,h+h_S}\right] \leq E\left[\eta^2_{t|h,h+h_L}\right] \qquad \text{for all } 0 \leq h_S < h_L. \tag{12}$$

In particular, setting $h = 1$, we have $\eta_{t|1,H} \equiv \hat{Y}^*_{t|t-1} - \hat{Y}^*_{t|t-H} = \sum_{j=1}^{H-1} \eta_{t|j,j+1}$, and

$$V\left[\eta_{t|1,H}\right] \geq V\left[\eta_{t|1,H-1}\right],$$

or, more generally,

$$V\left[\eta_{t|1,2}\right] \leq V\left[\eta_{t|1,3}\right] \leq \cdots \leq V\left[\eta_{t|1,H}\right]. \tag{13}$$

Again Wolak's (1987, 1989) testing framework can be applied here. Define the vector of mean-squared forecast revisions $\mathbf{\Delta}^\eta \equiv [\Delta^\eta_3, ..., \Delta^\eta_H]'$, where $\Delta^\eta_h \equiv E\left[\eta^2_{t|1,h}\right] - E\left[\eta^2_{t|1,h-1}\right] = E\left[\left(\hat{Y}^*_{t|t-1} - \hat{Y}^*_{t|t-h}\right)^2\right] - E\left[\left(\hat{Y}^*_{t|t-1} - \hat{Y}^*_{t|t-h+1}\right)^2\right]$. Then we can test the null hypothesis that the differences in mean-squared forecast revisions are weakly positive for all forecast horizons:

$$\begin{aligned} H_0 &: \boldsymbol{\Delta}^\eta \geq \mathbf{0}, \\ \text{vs. } H_1 &: \boldsymbol{\Delta}^\eta \in \mathbb{R}^{H-2}. \end{aligned} \tag{14}$$

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

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

"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 forecast revisions take the form

$$\begin{aligned} \eta_{t|1,2} &= \gamma\phi\varepsilon_{t-1} + (1-\gamma)\phi^2\varepsilon_{t-2} \\ \eta_{t|2,3} &= \gamma\phi^2\varepsilon_{t-2} + (1-\gamma)\phi^3\varepsilon_{t-3} \\ \eta_{t|1,3} &\equiv \eta_{t|1,2} + \eta_{t|2,3} = \gamma\phi\varepsilon_{t-1} + \phi^2\varepsilon_{t-2} + (1-\gamma)\phi^3\varepsilon_{t-3}, \end{aligned}$$

and so the variances of the one- and two-period forecast revisions are

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

We can then have a violation of the inequality in (13), i.e., $V(\eta_{t|1,2}) > V(\eta_{t|1,3})$, provided that

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

This condition is clearly not satisfied if $\gamma = 1$ (full optimality) or γ is close to one. Instead, it holds when γ is far away from one, representing either very sticky forecasts (e.g., $\gamma = 0.5$) or overshooting (e.g., $\gamma = 1.5$). This also includes the case where forecasters always delay their predictions by one period, $\gamma = 0$, and so report last period's optimal forecast.

2.7 Bounds on covariances of forecast revisions

Consider two forecast horizons, $h_L > h_S$, and the associated forecasts $\hat{Y}^*_{t|t-h_L}$ and $\hat{Y}^*_{t|t-h_S}$. Forecast optimality implies

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

It follows that $V\left[\eta_{t|h_S,h_L}\right] - 2Cov\left[Y_t, \eta_{t|h_S,h_L}\right] \leq 0$, or,

$$\begin{aligned}
V\left[\eta_{t|h_S,h_L}\right] &\leq 2Cov\left[Y_t, \eta_{t|h_S,h_L}\right] \\
E\left[\eta^2_{t|h_S,h_L}\right] &\leq 2E\left[Y_t \eta_{t|h_S,h_L}\right],
\end{aligned} \tag{15}$$

since $E\left[\eta_{t|h_S,h_L}\right] = 0$ under optimality. Hence we get the covariance bound

$$E\left[2Y_t \eta_{t|h_S,h_L} - \eta^2_{t|h_S,h_L}\right] \geq 0. \tag{16}$$

This limits the amount of variability in the forecast revisions, as a function of their covariance with the target variable. That is, monotonicity of the variance of the forecast errors implies an upper bound on the variance of the forecast revisions relative to the covariance of the revision with the target variable. This is a very intuitive result: if little relevant (in the sense of being correlated with the outcome) information arrives between the updating points, then the variance of the forecast revisions must be low. 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.

This can be tested by forming the vector $\mathbf{\Delta}^b \equiv \left[\Delta^b_2, ..., \Delta^b_H\right]'$, where $\Delta^b_h \equiv E\left[2Y_t \eta_{t|h-1,h} - \eta^2_{t|h-1,h}\right]$, for $h = 2, ..., H$ 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:

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

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:

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

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, $\tilde{Y}_{t|t-h} = \hat{Y}_{t|t-h_L} + u_{t|h}$, 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. We now 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.

Note that, under forecast optimality,

$$Y_t = \hat{Y}^*_{t|t-h} + e^*_{t|t-h},$$

with $E_{t-h}\left[e^*_{t|t-h}\right] = 0$, and also $\hat{Y}^*_{t|t-1} = \hat{Y}^*_{t|t-h} + \eta_{t|1,h}$, with $E_{t-h}\left[\eta_{t|1,h}\right] = 0$, for $h > 1$.

Essentially we can treat the short-horizon forecast, $\hat{Y}^*_{t|t-1}$, as a "proxy" for the target variable, in the sense that the monotonicity results involving the target variable *all* have equivalents that use the short-horizon forecast in its place. Importantly, unlike standard cases, the proxy in this case is *smoother than* the actual, rather than equal to the actual plus noise. This turns out to have important implications for the performance of optimality tests conducted on the short-horizon forecast in place of the actual outcome. The former may have better finite sample properties in situations where the measurement error is sizeable or the predictive R^2 of the forecasting model is low.

Using that $Cov\left[\hat{Y}^*_{t|t-h}, \eta_{t|1,h}\right] = 0$, we can show that the covariance of an arbitrary forecasts with the short-horizon forecast should be *decreasing* in the forecast horizon:

$$\begin{aligned} Cov\left[\hat{Y}^*_{t|t-h}, \hat{Y}^*_{t|t-1}\right] &= Cov\left[\hat{Y}^*_{t|t-h}, \hat{Y}^*_{t|t-h} + \eta_{t|1,h}\right] \\ &= V\left[\hat{Y}^*_{t|t-h}\right], \end{aligned}$$

which was shown earlier to be decreasing in the forecast horizon.

Similarly, we can also establish a bound on the variance of the forecast revision as a function of its covariance with the short-horizon forecast. Let $1 < h_S < h_L$. From (16) we have

$$\begin{aligned} V\left[\eta_{t|h_S,h_L}\right] &\leq 2Cov\left[Y_t, \eta_{t|h_S,h_L}\right] \\ &= 2Cov\left[\hat{Y}^*_{t|t-1} + e^*_{t|t-1}, \eta_{t|h_S,h_L}\right] \\ &= 2Cov\left[\hat{Y}^*_{t|t-1}, \eta_{t|h_S,h_L}\right], \end{aligned} \tag{17}$$

since $Cov\left[e^*_{t|t-1}, Z_{t-1}\right] = 0$. It follows that, as for (16),

$$E[2\hat{Y}^*_{t|t-1}\eta_{t|h_S,h_L} - \eta^2_{t|h_S,h_L}] \geq 0. \tag{18}$$

Thus the bound on the variance of the forecast revision can be expressed using only information in the forecasts, and it can be tested in the same way.

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. 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:

$$Y_t = \alpha_h + \beta_h \hat{Y}_{t|t-h} + u_{t|t-h}, \tag{19}$$

while forecast optimality can be tested through

$$\begin{aligned} H_0^h &: \alpha_h = 0 \cap \beta_h = 1 \\ H_1^h &: \alpha_h \neq 0 \cup \beta_h \neq 1. \end{aligned}$$

The MZ regression in (19) is usually applied separately to each forecast horizon. A simultaneous test of optimality across all horizons requires developing a different approach. We next present one such alternative.

3.1 Bonferroni bounds on MZ regressions

One approach, adopted in Capistrán (2007), is to run MZ regressions (19) for each horizon, $h = h_1, ..., h_H$. 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

Another approach is to stack the MZ equations for each horizon and estimate the model as a system:

$$\begin{bmatrix} Y_{t+h_1} \\ Y_{t+h_2} \\ Y_{t+h_3} \\ \vdots \\ Y_{t+h_H} \end{bmatrix} = \begin{bmatrix} \alpha_{h_1} \\ \alpha_{h_2} \\ \alpha_{h_3} \\ \vdots \\ \alpha_{h_H} \end{bmatrix} + \begin{bmatrix} \beta_{h_1} & 0 & 0 & \cdots & 0 \\ 0 & \beta_{h_2} & 0 & \cdots & 0 \\ 0 & 0 & \beta_{h_3} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \beta_{h_H} \end{bmatrix} \begin{bmatrix} \hat{Y}_{t+h_1|t} \\ \hat{Y}_{t+h_2|t} \\ \hat{Y}_{t+h_3|t} \\ \vdots \\ \hat{Y}_{t+h_H|t} \end{bmatrix} + \begin{bmatrix} e^1_{t+h_1} \\ e^2_{t+h_2} \\ e^3_{t+h_3} \\ \vdots \\ e^H_{t+h_H} \end{bmatrix}. \qquad (20)$$

The relevant hypothesis is now

$$\begin{aligned} H_0 &: \alpha_{h_1} = ... = \alpha_{h_H} = 0 \cap \beta_{h_1} = ... = \beta_{h_H} = 1 \qquad (21) \\ H_1 &: \alpha_{h_1} \neq 0 \cup ... \cup \alpha_{h_H} \neq 0 \cup \beta_{h_1} \neq 1 \cup ... \cup \beta_{h_H} \neq 1. \end{aligned}$$

For $h > 1$, the residuals in (20) 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 generalized MZ test

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, $\eta_{t|1,2}$,..., $\eta_{t|h_H-1,h_H}$. 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:

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

Rather than regressing the outcome variable on the one-period forecast, (19) suggests the following regression

$$Y_t = \alpha + \beta_{h_L} \hat{Y}_{t|t-h_H} + \sum_{h=h_1}^{h_H-1} \beta_h \eta_{t|h-1,h} + u_t. \qquad (22)$$

Under the null of optimality, the intercept in this regression should be zero and the slope coefficients should all be one. This reduces to the constraint that the target variable is equal to (one times) the short-horizon forecast, with zero intercept and zero weights on all longer-horizon forecasts:

$$\begin{aligned} H_0 &: \alpha = 0 \cap \beta_{h_1} = ...\beta_{h_H} = 1 \\ \text{vs. } H_1 &: \alpha \neq 0 \cup \beta_{h_1} \neq 1 \cup ... \cup \beta_{h_H} \neq 1. \end{aligned} \tag{23}$$

Notice that conducting the regression on forecast revisions rather than forecasts themselves reduces the extent of multicollinearity among the regressors and so could lead to a better-performing test in finite samples.

3.4 Regression tests without the target variable

All three of the above regression-based tests can be applied with the short-horizon forecast used in place of the target variable. In particular, a version of the MZ regression that is feasible when the target variable is *not* observed exploits the fact that, under optimality, we have, for $h_S < h_L$:

$$\begin{aligned} E_{t-h_L}\left[\hat{Y}^*_{t|t-h_S}\right] &= E_{t-h_L}\left[E_{t-h_S}\left[Y_t\right]\right] \\ &= E_{t-h_L}\left[Y_t\right] \\ &= \hat{Y}^*_{t|t-h_L}. \end{aligned}$$

Hence we can undertake a Mincer-Zarnowitz regression of the short-horizon forecast on a long-horizon forecast, and we should obtain the usual parameter restrictions:

$$\begin{aligned} \hat{Y}_{t|t-h_S} &= \alpha_{h_L,h_S} + \beta_{h_L,h_S}\hat{Y}_{t|t-h_L} + v_{t,h_L,h_S} \\ H_0^{h_L,h_S} &: \alpha_{h_L,h_S} = 0 \cap \beta_{h_L,h_S} = 1 \\ H_1^{h_L,h_S} &: \alpha_{h_L,h_S} \neq 0 \cup \beta_{h_L,h_S} \neq 1, \end{aligned} \tag{24}$$

for forecast horizons $h_1 \leq h_S < h_L \leq h_H$.

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 $\hat{Y}_{t|t-h_S} = E_{t-h_S}\left[Y_t\right]$ for all $h_S > 0$, then, trivially, $E_{t-h_S}\left[\hat{Y}_{t|t-h_S}\right] = E_{t-h_S}\left[Y_t\right]$, and so the forecast is an 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.

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

$$\begin{bmatrix} \hat{Y}_{t+h_2|t+1} \\ \vdots \\ \hat{Y}_{t+h_H|t+h_H-1} \end{bmatrix} = \begin{bmatrix} \tilde{\alpha}_{h_2} \\ \vdots \\ \tilde{\alpha}_{h_H} \end{bmatrix} + \begin{bmatrix} \tilde{\beta}_{h_2} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \tilde{\beta}_{h_H} \end{bmatrix} \begin{bmatrix} \hat{Y}_{t+h_2|t} \\ \vdots \\ \hat{Y}_{t+h_H|t} \end{bmatrix} + \begin{bmatrix} \tilde{u}^{h_2}_{t+1} \\ \vdots \\ \tilde{u}^{h_H}_{t+h_H} \end{bmatrix}, \qquad (25)$$

with the hypothesis

$$H_0 : \alpha_{h_2} = ... = \alpha_{h_H} = 0 \cap \beta_{h_2} = ... = \beta_{h_H} = 1.$$

Finally, (22) suggests running a regression of the short-run forecast on an intercept, the long-run forecast and the sequence of all but one of the interim forecast revisions:

$$\hat{Y}_{t|t-h_1} = \alpha + \beta_{h_H} \hat{Y}_{t|t-h_H} + \sum_{h=h_2}^{h_H-1} \beta_h \eta_{t|h-1,h} + u_t, \qquad (26)$$

where $\eta_{t|h_S,h_L} \equiv \hat{Y}_{t|t-h_S} - \hat{Y}_{t|t-h_L}$ for $h_S < h_L$. Rationality now implies

$$\begin{aligned} H_0 &: \alpha = 0 \cap \beta_{h_2} = ... = \beta_{h_H} = 1 \\ H_1 &: \alpha \neq 0 \cup \beta_{h_1} \neq 1 \cup ... \cup \beta_{h_H} \neq 1. \end{aligned}$$

Note that we do *not* include the forecast revision between h_1 and h_2 on the right-hand side above. The right-hand side only uses forecasts with horizons *longer than* the forecast horizon of the variable being used to proxy for the target variable.

4 Stationarity and Tests of Forecast Optimality

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

$$\tilde{Y}_t = \sum_{i=0}^{t} \theta_i \varepsilon_{t-i}, \qquad (27)$$

where $\varepsilon_{t-i} \sim WN(0, 1)$ is serially uncorrelated white noise and $\lim_{t\to\infty} \sum_{i=0}^{t} \theta_i^2 < \infty$. 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

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

Then, by the law of iterated expectations,

$$E[(Y_{t+h} - \hat{Y}^*_{t+h|t-j})^2] \geq E[(Y_{t+h} - \hat{Y}^*_{t+h|t})^2] \quad \text{for } j \geq 1. \tag{29}$$

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 (29) that

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

The covariance stationarity assumption is clearly important here. (30) does not follow from (29) if, say, there is a deterministic reduction in the variance of Y between periods $t+h$ and $t+h+j$. Suppose for example that

$$Y_\tau = \begin{cases} \mu + \sigma\varepsilon_\tau & \text{for } \tau \leq t+h \\ \mu + \frac{\sigma}{2}\varepsilon_\tau & \text{for } \tau > t+h \end{cases}, \tag{31}$$

where ε_τ is zero-mean white noise. This could be a stylized example of the "Great Moderation". Clearly (30) is now violated as $\hat{Y}^*_{t+h+j|t} = \hat{Y}^*_{t+h|t} = \mu$, and so[5]

$$E[(Y_{t+h+j} - \hat{Y}^*_{t+h+j|t})^2] = \frac{\sigma^2}{4} < \sigma^2 = E[(Y_{t+h} - \hat{Y}^*_{t+h|t})^2] \quad \text{for } j \geq 1. \tag{32}$$

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).

[5] Notice here that the expectation, $E[.]$, 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 (31) 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 (29) rather than (30). 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,

$$E_{t-h_S}[(Y_t - \hat{Y}^*_{t|t-h_L})^2] \geq E_{t-h_S}[(Y_t - \hat{Y}^*_{t|t-h_S})^2] \quad \text{for } h_L \geq h_S. \tag{33}$$

Moreover, by the law of iterated expectations,

$$E[(Y_t - \hat{Y}^*_{t|t-h_L})^2] \geq E[(Y_t - \hat{Y}^*_{t|t-h_S})^2] \quad \text{for } h_L \geq h_S. \tag{34}$$

This result is quite robust. For example, with a break in the variance, (31), we have $\hat{Y}^*_{t|t-h_L} = \hat{Y}^*_{t|t-h_S} = \mu$, and

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

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:

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

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 will show that the variance bound results pertain to a more general class of stochastic processes that do not require covariance stationarity. These processes take the form

$$Y_t = f(t, \theta) + Y_0 + \sum_{i=0}^{t} \theta_{it} \varepsilon_{t-i}. \tag{35}$$

Variance bound results for fixed-event forecasts can be established for this family of processes under the following assumption:

Assumption 1 *For all t, θ_{it} is a sequence of deterministic coefficients such that,* $\lim_{I \to \infty} \sum_{i=0}^{I} \theta_{it}^2 < \infty$.

Hence θ_{it} are deterministic weights on past shocks which depend both on the current time, t, and the "lag", 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 (31), we get

$$\theta_{0,\tau} = \begin{cases} \sigma & \text{for } \tau \leq t + h \\ \frac{\sigma}{2} & \text{for } \tau > t + h \end{cases}, \tag{36}$$

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

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.[6] For these reasons it is important to shed light on the finite sample performance of the

[6]One 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.

various forecast optimality tests. We do so in this section by means of a Monte Carlo simulation study. We first describe the simulation design and then present the size and power results.

5.1 Simulation setup

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

$$\begin{aligned} Y_t &= \mu_y + \phi\left(Y_{t-1} - \mu_y\right) + \varepsilon_t, \;\; t = 1, 2, ..., T = 100 \\ \varepsilon_t &\sim iid\; N\left(0, \sigma_\varepsilon^2\right). \end{aligned} \tag{37}$$

We calibrate the parameters to quarterly US CPI inflation data:

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

Optimal forecasts are formed as the conditional expectation:

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

for $h = 1, 2, .., H$ and $H \in \{\ 4\ ,\ 8\ \}$.

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

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

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);[7] and (iii) *high*, $\sigma_\psi/\sigma_y = 1$ (roughly twice that for GDP growth).

[7] The "medium" value is calibrated to match US GDP growth data, as reported by Faust, Rogers and Wright (2005).

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:

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

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 magnitude is *increasing* in the horizon:

$$\sigma_{\xi,h} = \frac{2(h-1)}{H-1} \times 0.65\sigma_y, \;\; \text{for } h = 1, 2, ..., H.$$

Forecasts affected by noise whose magnitude is *decreasing* in the horizon take the form:

$$\sigma_{\xi,h} = \frac{2(H-h)}{H-1} \times 0.65\sigma_y, \;\; \text{for } h = 1, 2, ..., H.$$

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

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

To capture "sticky" forecasts we set $\gamma = 1/2$, whereas "overshooting" forecasts have $\gamma = 1.5$.

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.

5.2 Results

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.[8] The vector MZ test is also hugely oversized, while the size of the univariate multi-horizon MZ regression test is close to the nominal value of 10%. Because

[8] Conventionally, 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.

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 or MSFR bounds have much power to detect deviations from forecast optimality. This holds across all three levels of measurement error. In contrast, the covariance bound 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.[9] 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 MZ regression on all revisions (22) 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 (8). 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. Interestingly, the power is even stronger for the covariance bound test conducted on the short-horizon forecast (18), although in the presence of large measurement errors the similar bound conducted on the actual value (16) clearly has weaker power than the bound based on decreasing MSF-values. The MSE and MSFR bounds have zero power for this type of deviation from forecast optimality. The univariate MZ regression based on all revisions (22) performs worse than in the case with horizon-independent noise, although this test still has power of 50%-60% when the dependent variable is the smoother short-horizon forecast.[10]

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

[9] The covariance bound (16) works so well because noise in the forecast increases $E[\eta^2_{t|h_S,h_L}]$ without affecting $E[Y_t\eta_{t|h_S,h_L}]$, thereby making it less likely that $E[Y_t\eta_{t|h_S,h_L} - \eta^2_{t|h_S,h_L}] \geq 0$ holds.

[10] For 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.

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 joint MZ regression on all revisions (22) have superior power.

The univariate joint MZ test (22) 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 final scenario assumes sticky updating in the forecasts. In this case, shown in Panel D of Table 2, only the univariate joint MZ regression (22) 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 joint MZ regression is based on the short-term forecasts only.

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. In all cases we find that the size of the tests falls well below the nominal size, 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. Even better performance is, however, found for the univariate MZ regression on all revisions, 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 all four types of deviations from forecast efficiency. Bonferroni bound tests conducted on the regression and second moment tests also perform quite well when they include the best 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.[11] We use quarterly observations 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., $h = 0, 1, 2, 3, 4, 5$. Some observations are missing and so we have a total of just under 80 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 joint MZ regression that uses the short-run forecast as the target variable, (26), and a mild violation of the increasing mean-squared forecast revision test, (13).

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 mean squared forecast, and the univariate joint MZ test on revisions all lead to rejections. Figure 1 illustrates the rejection of the variance bound based on the forecasts and shows that, in contradiction with (8) 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, (13), and a rejection through the univariate joint MZ 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 ortogonal 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.

[11] We are grateful to Jonathan Wright for providing the data.

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 (16), (18) and the univariate MZ regressions on revisions and the long-run forecast (22), (26). 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 Gourieroux 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.

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] 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.

[14] 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.

[15] 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.

[16] Newey, Whitney K., and Kenneth D. West, 1987, A Simple, Positive Semidefinite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix, Econometrica, 55, 703-708.

[17] Nordhaus, William D., 1987, Forecasting Efficiency: Concepts and Applications. Review of Economics and Statistics 69, 667-674.

[18] Patton, Andrew J., and Allan Timmermann, 2007, Testing Forecast Optimality under Unknown Loss. Journal of the American Statistical Association 102, 1172-1184.

[19] Patton, Andrew J., and Allan Timmermann, 2008, Predictability of output growth and inflation: A multi-horizon survey approach. Unpublished manuscript, Duke and UCSD.

[20] 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.

[21] Schmidt, Peter, 1974, The Asymptotic Distribution of Forecasts in the Dynamic Simulation of an Econometric Model. Econometrica 42, 303-309.

[22] 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.

[23] Wolak, Frank A., 1989, Testing Inequality Constraints in Linear Econometric Models. Journal of Econometrics 31, 205-235.

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

	H = 4			**H = 8**		
Meas. error variance:	*High*	*Med*	*Zero*	*High*	*Med*	*Zero*
SIZE						
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
MZ on all revisions	11.3	11.5	11.0	12.4	11.8	11.0
MZ on all revisions, 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 $\mu_y = 0.75$. 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) tests and use simulated critical values based on a mixture of chi-squared variables.

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

	H = 4			H = 8		
Meas. error variance:	*High*	*Med*	*Zero*	*High*	*Med*	*Zero*
PANEL A: Equal noise across all forecast horizons						
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
MZ on all revisions	91.9	98.1	99.6	85.3	95.9	99.0
MZ on all revisions, 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 $\mu_y = 0.75$. 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%.

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

	H = 4			H = 8		
Meas. error variance:	*High*	*Med*	*Zero*	*High*	*Med*	*Zero*
PANEL B: Noise increases with the horizon						
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
MZ on all revisions	11.7	12.3	11.9	13.1	13.6	12.9
MZ on all revisions, 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

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

	H = 4			H = 8		
Meas. error variance:	*High*	*Med*	*Zero*	*High*	*Med*	*Zero*
PANEL C: Noise decreases with the horizon						
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
MZ on all revisions	100.0	100.0	100.0	100.0	100.0	100.0
MZ on all revisions, 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

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

	H = 4			H = 8		
Meas. error variance:	*High*	*Med*	*Zero*	*High*	*Med*	*Zero*
			PANEL D: Sticky updating			
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
MZ on all revisions	33.2	44.8	59.1	27.5	35.2	49.9
MZ on all revisions, 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

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

	H = 4			H = 8		
Meas. error variance:	*High*	*Med*	*Zero*	*High*	*Med*	*Zero*
			PANEL E: Over-shooting (TBC)			
Inc MSE						
Dec COV						
COV bound						
Dec MSF						
Inc MSFR						
Dec COV, h=1						
COV bound, h=1						
Inc MSE & Dec MSF						
Inc MSE & Inc MSFR						
MZ on all revisions						
MZ on all revisions, h=1						
Bonf, using actuals						
Bonf, using forecasts only						
Bonf, all tests						

Table 3: Forecast optimality tests for Greenbook forecasts

Series:	*Growth*	*Deflator*	*Inflation*
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
MZ on all revisions	0.709	0.000*	0.001*
MZ on all revisions, 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) tests and use simulated critical values based on a mixture of chi-squared variables.

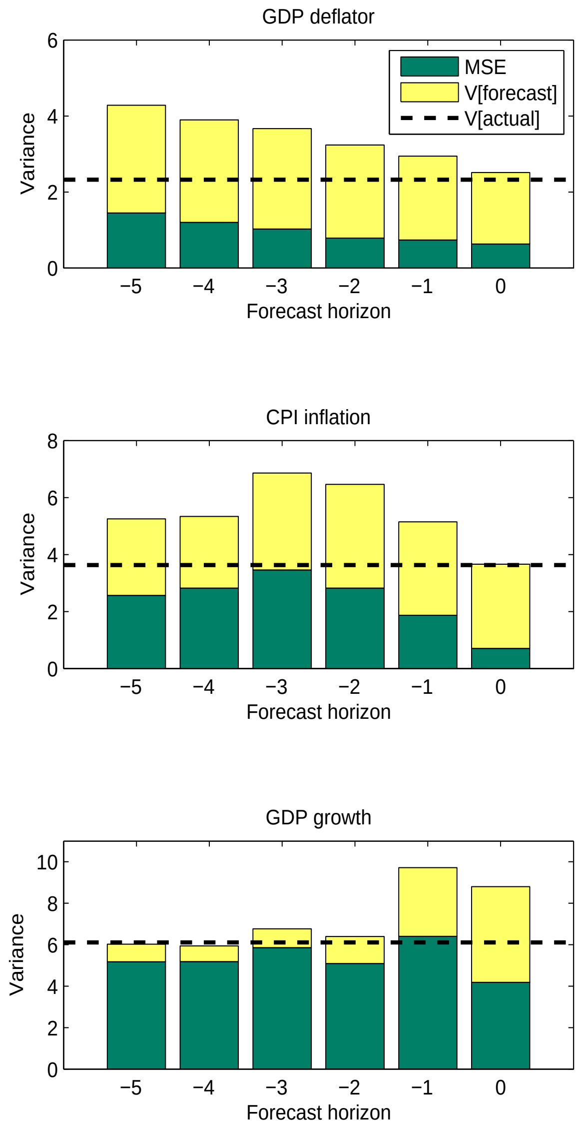

Figure 1: *Mean squared errors and forecast variances, for US GDP deflator, CPI inflation and GDP growth.*

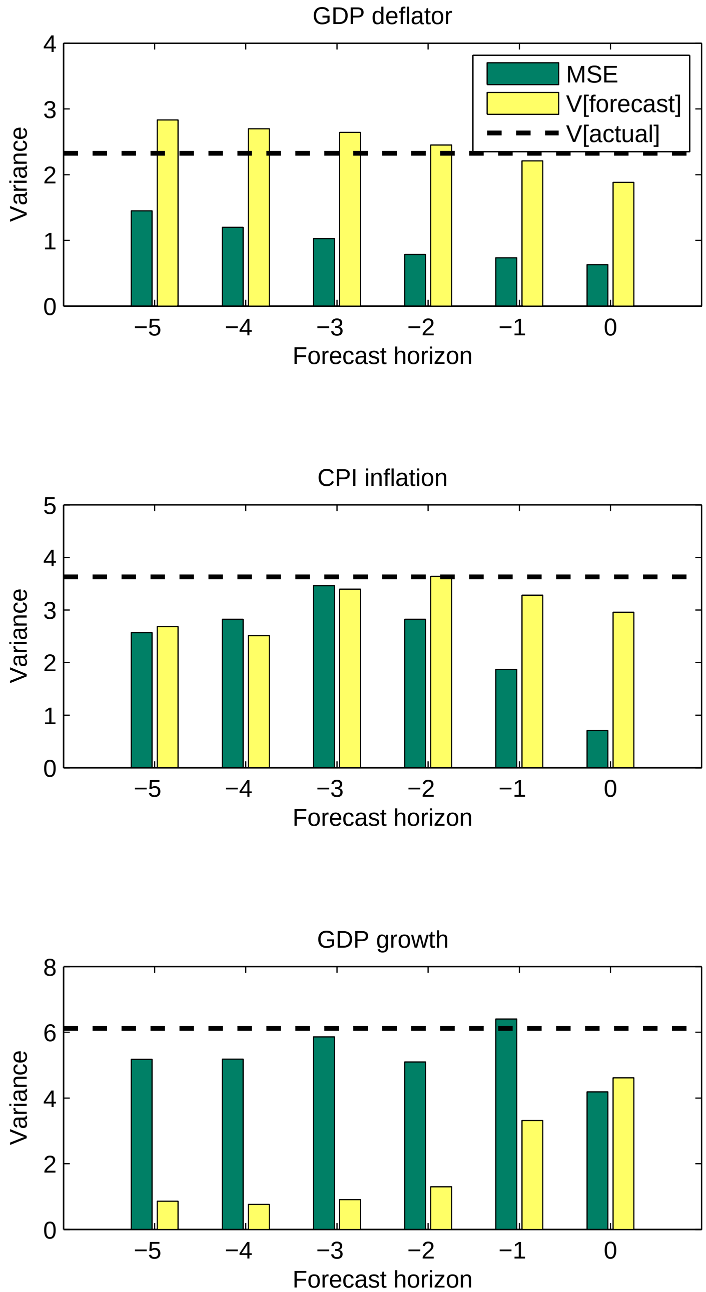

Figure 2: *Mean squared errors and forecast variances, for US GDP deflator, CPI inflation and GDP growth.*

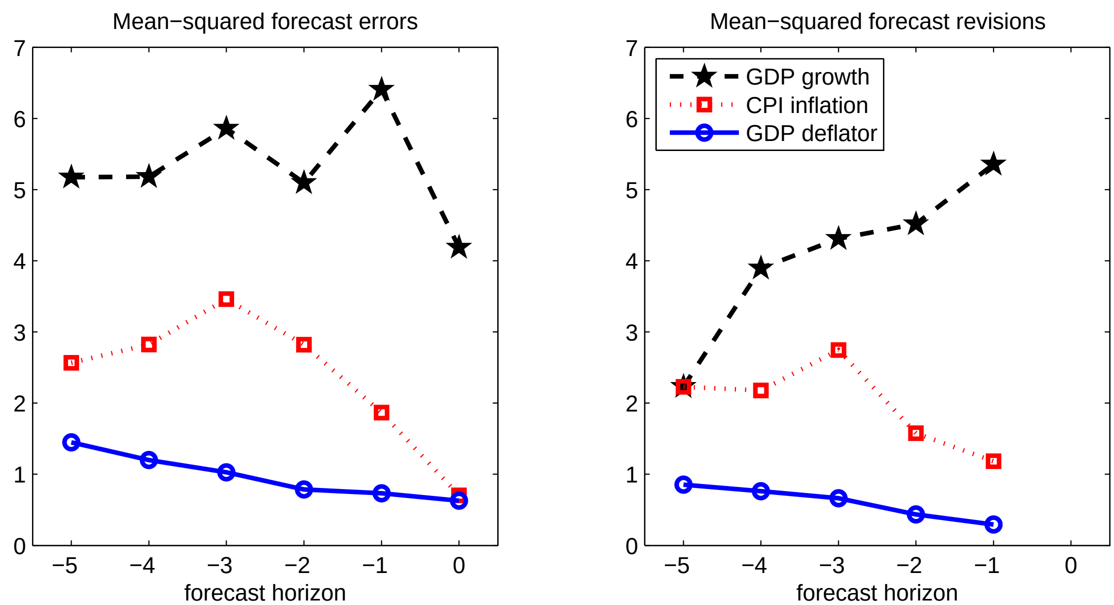

Figure 3: *Mean squared errors (left panel) and mean-squared forecast revisions (right panel), for US GDP deflator, CPI inflation and GDP growth.*