DISCUSSION PAPER SERIES

No. 8194

FORECAST RATIONALITY TESTS
BASED ON MULTI-HORIZON BOUNDS

Andrew J Patton and Allan G Timmermann

FINANCIAL ECONOMICS

Centre for Economic Policy Research

www.cepr.org

FORECAST RATIONALITY TESTS BASED ON MULTI-HORIZON BOUNDS

Andrew J Patton, Duke University
Allan G Timmermann, University of California, San Diego

ABSTRACT

Forecast Rationality Tests Based on Multi-Horizon Bounds*

Forecast rationality under squared error loss implies various bounds on second moments of the data across forecast horizons. For example, the mean squared forecast error should be increasing in the horizon, and the mean squared forecast should be decreasing in the horizon. We propose rationality tests based on these restrictions, including new ones that can be conducted without data on the target variable, and implement them via tests of inequality constraints in a regression framework. A new optimal revision test based on a regression of the target variable on the long-horizon forecast and the sequence of interim forecast revisions is also proposed. 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 presented.

Andrew J Patton
Department of Economics
Duke University
213 Social Sciences Building
Durham, NC 27708-0097
USA

Email: andrew.patton@duke.edu

Allan G Timmermann
University of California San Diego
Rady School of Management
9500 Gilman Drive
La Jolla CA 92093-0553
USA

Email: atimmerm@ucsd.edu

1   Introduction

Forecasts recorded at multiple horizons, e.g., from one to several quarters into the future, are commonly reported in empirical work. 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 multiple horizons. Similarly, econometric models are commonly used to generate multi-horizon forecasts, see, e.g., Clements (1997), Faust and Wright (2009) and Marcellino, Stock and Watson (2006). With the availability of such multi-horizon forecasts, there is a growing need for tests of optimality that 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 rationality and compares tests based on these implications with extant methods.

A well-known implication of forecast rationality is that, under squared error loss, the mean squared forecast error should be a weakly increasing 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 weakly decreasing function of the forecast horizon. Intuitively, this property holds because the variance of the expectation conditional on a large information set (corresponding to a short forecast 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 contraints using the methods of Gourieroux et al. (1982) and Wolak (1987, 1989), and the bootstrap methods of White (2000) and Hansen (2005).

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 (2009), revisions to macroeconomic variables can be very considerable and so 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, a complete test that imposes internal consistency restrictions on the forecast revisions is shown to give rise to a univariate optimal revision 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. This can substantially increase the power of the test.

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), Davies

and Lahiri (1995), Clements (1997), Isiklar et al. (2006), and Patton and Timmermann (2008, 2010b). 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 and other forms of data heterogeneity.

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 univariate optimal revision test 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 the Federal Reserve's 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 variance bound implications of optimality of forecasts across multiple horizons and the associated tests. Section 3 considers regression-based tests of forecast optimality and Section 4 presents some extensions of our main results to cover data heterogeneity and heterogeneity in the forecast horizons. Section 5 presents results from a Monte Carlo study, while Section 6 provides an empirical application to Federal Reserve Greenbook forecasts. Section 7 concludes.

2 Multi-Horizon Bounds and 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 collected in Appendix A.

2.1 Assumptions and background

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 are assumed to be conditioned on the information set available at time $t - h$, $\mathcal{F}_{t-h}$, which is taken to be the filtration of σ-algebras generated by $\left\{ \tilde{Z}_{t-h-k}; k \geq 0 \right\}$, where $\tilde{Z}_{t-h}$ is a vector of predictor variables. This need not (only) comprise past and current values of Y. 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 ($h_L > h_S$). 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 will develop a variety of forecast optimality tests based on corollaries to Theorem 1 below. In so doing, we 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 finite samples (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) and use inadmissable forecasts based on plug-in estimators. In practice forecasts from surveys and forecasts reported by central banks reflect considerable judgmental information which is difficult to handle using the methods developed by West (1996) and West and

McCracken (1998).

The "real time" macroeconomics literature has demonstrated the presence of large and prevalent measurement errors affecting a variety of macroeconomic variables, see Corradi, Fernandez and Swanson (2009), Croushore (2006), Croushore and Stark (2001), Diebold and Rudebusch (1991), and Faust, Rogers and Wright (2005). In such situations it is useful to have tests that do not require data on the target variable and we present such tests below. These tests 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 has 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.

Under squared error loss, we have the following well-known theorem (see, e.g., Granger (1969)):

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

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

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

Section 4.3 shows that this result can be extended to a broader class of loss functions, but for now we keep the more familiar assumption of quadratic loss. 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.

6

2.2 Monotonicity of mean squared errors and forecast revisions

From forecast optimality under squared-error loss 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:

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

In situations where the predicted variable is not observed (or only observed with error), one can instead compare medium- and long-horizon forecasts with the short-horizon forecast. Define a forecast revision as

$$d^*_{t|h_S,h_L} \equiv \hat{Y}^*_{t|t-h_S} - \hat{Y}^*_{t|t-h_L} \quad \text{for } h_S < h_L.$$

The corollary below shows that the bounds on mean squared forecast errors that follow immediately from forecast rationality under squared-error loss also apply to mean squared forecast *revisions.*

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

$$(a) \quad E\left[e^{*2}_{t|t-h_S}\right] \leq E\left[e^{*2}_{t|t-h_L}\right] \quad \text{for } h_S < h_L, \tag{2}$$

and

$$(b) \quad E\left[d^{*2}_{t|h_S,h_M}\right] \leq E\left[d^{*2}_{t|h_S,h_L}\right] \quad \text{for } h_S < h_M < h_L. \tag{3}$$

The inequalities are strict if more forecast-relevant information becomes available as the forecast horizon shrinks to zero, see, e.g., Diebold (2001) and Patton and Timmermann (2007a).

2.3 Testing monotonicity in squared forecast errors and forecast revisions

Corollary 1 suggests testing forecast optimality via a test of the weak monotonicity in the "term structure" of mean squared errors, Eq.(2), to use the terminology of Patton and Timmermann (2008). This feature of rational forecasts is relatively widely known, but has, with the exception

7

of Capistran (2007), generally not been used to test forecast optimality. Capistran's 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 Eq. (2).

We consider ranking the MSE-values for a set of forecast horizons $h = h_1, h_2, ..., h_H$. Denoting the population value of the MSEs by $\boldsymbol{\mu}^e = [\mu_1^e, ..., \mu_H^e]'$, with $\mu_j^e \equiv E[e_{t|t-h_j}^2]$, and defining the associated MSE differentials as $\Delta_j^e \equiv \mu_j - \mu_{j-1} = E\left[e_{t|t-h_j}^2\right] - E\left[e_{t|t-h_{j-1}}^2\right]$, we can rewrite the inequalities in (2) as

$$\Delta_j^e \geq 0, \quad \text{for } j = 2, ..., H. \tag{4}$$

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:

$$H_0 : \boldsymbol{\Delta}^e \geq \mathbf{0} \quad \text{vs.} \quad H_1 : \boldsymbol{\Delta}^e \ngeq \mathbf{0}, \tag{5}$$

where the $(H-1) \times 1$ vector of MSE-differentials is given by $\boldsymbol{\Delta}^e \equiv [\Delta_2^e, ..., \Delta_H^e]'$. As in Patton and Timmermann (2010a), 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, $\sum_{i=0}^{H-1} \omega(H-1, i)\chi^2(i)$, where $\omega(H-1, i)$ are the weights and $\chi^2(i)$ is a chi-squared variable with i degrees of freedom. The key computational difficulty in implementing this test is obtaining the weights. These weights equal the probability that the vector $Z \sim N(\mathbf{0}, \Sigma)$ has exactly i positive elements, where Σ is the long-run covariance matrix of the estimated parameter vector, $\hat{\boldsymbol{\Delta}}^e$. One straightforward way to estimate these weights is via simulation, see Wolak (1989, p. 215). An alternative is to compute these weights in closed form, using the work of Kudo (1963) and Sun (1988), which is faster when the dimension is not too large (less than 10). (We thank Raymond Kan for suggesting this alternative approach to us, and for generously providing Matlab code to implement this approach.) When the dimension is large, one can alternatively use the bootstrap methods in White (2000) and Hansen (2005), which are explicitly designed to work for high-dimension problems.

Note that the inequality in Eq. (2) implies a total of $H(H-1)/2$ pair-wise inequalities, not just the $H-1$ inequalities obtained by comparing "adjacent" forecast horizons. In a related testing problem, Patton and Timmermann (2010a) consider tests based both the complete set of inequalities and the set of inequalities based only on "adjacent" horizons (portfolios, in their case) and found little different in size or power of these two approaches. For simplicity, we consider only inequalities based on "adjacent" horizons.

Wolak's testing framework can also be applied to the bound on the mean squared forecast revisions (MSFR). To this end, define the $(H-2) \times 1$ vector of mean-squared forecast revisions $\boldsymbol{\Delta}^d \equiv \left[\Delta_3^d, ..., \Delta_H^d\right]'$, where $\Delta_j^d \equiv E\left[d_{t|h_1,h_j}^2\right] - E\left[d_{t|h_1,h_{j-1}}^2\right]$. Then we can test the null hypothesis that differences in mean-squared forecast revisions are weakly positive for all forecast horizons:

$$H_0 : \boldsymbol{\Delta}^d \geq \mathbf{0} \quad \text{vs.} \quad H_1 : \boldsymbol{\Delta}^d \ngeq \mathbf{0}. \tag{6}$$

2.4 Monotonicity of mean squared *forecasts*

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

Corollary 2 *Under the assumptions of Theorem 1 and S1, we have*

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

This result is closely related to Corollary 1 since $V[Y_t] = V\left[\hat{Y}_{t|t-h}^*\right] + E\left[e_{t|t-h}^{*2}\right]$. A weakly increasing pattern in MSE directly implies a weakly *decreasing* pattern in the variance of the forecasts. Hence, one aspect of forecast optimality can be tested *without* the need for a measure of the target variable. Notice again that since $E\left[\hat{Y}_{t|t-h}^*\right] = E[Y_t]$ for all h, we obtain the following inequality on the mean-squared forecasts:

$$E\left[\hat{Y}_{t|t-h_S}^{*2}\right] \geq E\left[\hat{Y}_{t|t-h_L}^{*2}\right] \quad \text{for any } h_S < h_L. \tag{7}$$

9

A test of this implication can again be based on Wolak's (1989) approach by defining the vector $\boldsymbol{\Delta}^f \equiv \left[\Delta_2^f, ..., \Delta_H^f \right]'$, where $\Delta_j^f \equiv E \left[\hat{Y}_{t|t-h_j}^{*2} \right] - E \left[\hat{Y}_{t|t-h_{j-1}}^{*2} \right]$ and testing the null hypothesis that differences in mean squared forecasts (MSF) are weakly negative for all forecast horizons:

$$H_0 : \boldsymbol{\Delta}^f \leq \mathbf{0} \quad \text{vs.} \quad H_1 : \boldsymbol{\Delta}^f \nleq \mathbf{0}. \tag{8}$$

It is worth pointing out a limitation 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, and not direct tests of forecast rationality, see the discussion in Pesaran and Weale (2006). For example, forecasts from 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.

2.5 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

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

Similarly, the covariance of the short-term forecast with another forecast should be decreasing in the other forecast's horizon:

$$Cov \left[\hat{Y}_{t|t-h_L}^*, \hat{Y}_{t|t-h_S}^* \right] = Cov \left[\hat{Y}_{t|t-h_L}^*, \hat{Y}_{t|t-h_L}^* + d_{t|h_S,h_L}^* \right] = V \left[\hat{Y}_{t|t-h_L}^* \right].$$

Thus we obtain the following:

Corollary 3 *Under the assumptions of Theorem 1 and S1, we have, for any $h_S < h_L$,*

$$(a) \quad Cov \left[\hat{Y}_{t|t-h_S}^*, Y_t \right] \geq Cov \left[\hat{Y}_{t|t-h_L}^*, Y_t \right]$$

Moreover, for any $h_S < h_M < h_L$,

$$(b) \quad Cov \left[\hat{Y}_{t|t-h_M}^*, \hat{Y}_{t|t-h_S}^* \right] \geq Cov \left[\hat{Y}_{t|t-h_L}^*, \hat{Y}_{t|t-h_S}^* \right].$$

10

Once again, using $E\left[\hat{Y}^*_{t|t-h}\right] = E\left[Y_t\right]$, it follows that we can express the above bounds as simple expectations of products:

$$E\left[\hat{Y}^*_{t|t-h_S} Y_t\right] \geq E\left[\hat{Y}^*_{t|t-h_L} Y_t\right]$$

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

As for the above cases, these implications can again be tested by defining the vector $\boldsymbol{\Delta}^c \equiv [\Delta^c_2, ..., \Delta^c_H]'$, where $\Delta^c_j \equiv E\left[\hat{Y}^*_{t|t-h_j} Y_t\right] - E\left[\hat{Y}^*_{t|t-h_{j-1}} Y_t\right]$ and testing:

$$H_0 : \boldsymbol{\Delta}^c \leq \mathbf{0} \quad \text{vs.} \quad H_1 : \boldsymbol{\Delta}^c \nleq \mathbf{0} \tag{9}$$

using Wolak's (1989) approach.

2.6 Bounds on covariances of forecast revisions

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

$$(a) \quad V\left[d^*_{t|h_S,h_L}\right] \leq 2Cov\left[Y_t, d^*_{t|h_S,h_L}\right] \quad \text{for any } h_S < h_L. \tag{10}$$

Moreover,

$$(b) \quad V\left[d^*_{t|h_M,h_L}\right] \leq 2Cov\left[\hat{Y}^*_{t|t-h_S}, d^*_{t|h_M,h_L}\right] \quad \text{for any } h_S < h_M < h_L. \tag{11}$$

For testing purposes, using $E\left[d^*_{t|h_S,h_L}\right] = 0$, we can use the more convenient inequalities:

$$E\left[d^{*2}_{t|h_S,h_L}\right] \leq 2E\left[Y_t d^*_{t|h_S,h_L}\right], \text{ for any } h_S < h_L \quad \text{or} \tag{12}$$

$$E\left[d^{*2}_{t|h_M,h_L}\right] \leq 2E\left[\hat{Y}^*_{t|t-h_S} d^*_{t|h_M,h_L}\right] \quad \text{for any } h_S < h_M < h_L.$$

11

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

$$H_0 : \boldsymbol{\Delta}^b \geq \boldsymbol{0} \quad \text{vs.} \quad H_1 : \boldsymbol{\Delta}^b \ngeq \boldsymbol{0}.$$

2.7 Monotonicity of covariances with the forecast error

We finally consider bounds on the covariance of the forecast error with the target variable or with the forecast revision, and corresponding versions of these results that may be implemented when data on the target variable are not available or not reliable. These bounds are perhaps less intuitive than the earlier ones and so will not be further pursued, but are included for completeness.

Corollary 5 *Under the assumptions of Theorem 1 and S1, we have,*

(a) for any $h_S < h_L$,

$$Cov\left[e_{t|t-h_S}^*, Y_t\right] \leq Cov\left[e_{t|t-h_L}^*, Y_t\right].$$

(b) for any $h_S < h_M < h_L$,

$$Cov\left[d_{t|h_S,h_M}^*, \hat{Y}_{t|t-h_S}^*\right] \leq Cov\left[d_{t|h_S,h_L}^*, \hat{Y}_{t|t-h_S}^*\right],$$

and (c)

$$Cov\left[e_{t|t-h_M}^*, d_{t|h_S,h_M}^*\right] \leq Cov\left[e_{t|t-h_L}^*, d_{t|h_S,h_L}^*\right].$$

The first part follows from the simple intuition that as the forecast horizon grows, the forecast explains less and less of the target variable, and thus the forecast error becomes more and more like the target variable. The last inequality links the forecast error made a time $t - h$ to the forecast revision made between time $t - h$ and some shorter horizon. This result may prove useful where

12

the target variable (and thus the forecasts) is very persistent, as the variables in this bound are differences between actuals and forecasts, or between forecasts, and will be less persistent than the original variables. In applications where the target variable is not available a corresponding result involves using the short horizon forecast in place of the target variable. Doing so gives a result for the variance of the forecast revision, which was already presented above in Corollary 1, and so it is not repeated here.

2.8   Summary of test methods

The tests presented here are based on statistical properties of either the outcome variable, Y_t, the forecast error, $e^*_{t|t-h}$, the forecast, $\hat{Y}^*_{t|t-h}$, or the forecast revision, $d^*_{t|h_S,h_L}$. The table below shows that the tests discussed so far provide an exhaustive list of all possible bounds tests based on these four variables and their mutual relations. The table lists results as the forecast horizon increases ($h \uparrow$), and for the forecast revision relations we keep the short horizon (h_S) fixed:

	Y_t	$e^*_{t|t-h}$	$\hat{Y}^*_{t|t-h}$	$d^*_{t|h_S,h_L}$
Y_t	σ_y^2	Cov $\uparrow$	Cov $\downarrow$	Cov bound
$e^*_{t|t-h}$		MSE $\uparrow$	Cov=0	Cov $\uparrow$
$\hat{Y}^*_{t|t-h}$			MSF $\downarrow$	Cov $\uparrow$
$d^*_{t|h_S,h_L}$				MSFR $\uparrow$

Almost all existing optimality tests focus on cell (2,3), i.e., that forecast errors are uncorrelated with the forecast, which is what conventional rationality regressions effectively test as can be seen by subtracting the forecast from both sides of the regression. Capistran (2007) studies the increasing MSE property, cell (2,2). Our analysis generalizes extant tests to the remaining elements. We pay particular attention to cells (3,3) (3,4) and (4,4), which do not require data on the target variable, and thus may of use when this variable is measured with error or unavailable. Appendix B presents an analytical illustration of the bounds established in this section for an AR(1) process.

13

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 MZ regression takes the form:

$$Y_t = \alpha_h + \beta_h \hat{Y}_{t|t-h} + v_{t|t-h}. \tag{13}$$

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

Corollary 6 *Under the assumptions of Theorem 1 and S1, the population values of the parameters in the Mincer-Zarnowitz regression in Eq. (13) satisfy*

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

The MZ regression in Eq. (13) 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 (13) for each horizon, $h = h_1, ..., h_H$ and 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. Forecast optimality is rejected 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 bound approach is to stack the MZ equations for each horizon and estimate them as a system:

$$
\begin{bmatrix} Y_{t+h_1} \\ \vdots \\ Y_{t+h_H} \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_H \end{bmatrix} + \begin{bmatrix} \beta_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \beta_H \end{bmatrix} \begin{bmatrix} \hat{Y}_{t+h_1|t} \\ \vdots \\ \hat{Y}_{t+h_H|t} \end{bmatrix} + \begin{bmatrix} v_{t+h_1|t} \\ \vdots \\ v_{t+h_H|t} \end{bmatrix}. \tag{14}
$$

The relevant hypothesis is now

$$
H_0 \quad : \quad \alpha_1 = ... = \alpha_H = 0 \cap \beta_1 = ... = \beta_H = 1 \tag{15}
$$

$$
\text{vs.} \quad H_1 \quad : \quad \alpha_1 \neq 0 \cup ... \cup \alpha_H \neq 0 \cup \beta_1 \neq 1 \cup ... \cup \beta_H \neq 1.
$$

For $h > 1$, the residuals in Eq. (14) will, even under the null, 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 a *univariate* regression. 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, $d_{t|h_1,h_2}, ..., d_{t|h_{H-1},h_H}$. 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:

$$
\hat{Y}_{t|t-h_1} \equiv \hat{Y}_{t|t-h_H} + \sum_{j=1}^{H-1} d_{t|h_j,h_{j+1}}. \tag{16}
$$

Rather than regressing the outcome variable on the one-period forecast, we propose the following "optimal revision" regression:

$$
Y_t = \alpha + \beta_H \hat{Y}_{t|t-h_H} + \sum_{j=1}^{H-1} \beta_j d_{t|h_j,h_{j+1}} + u_t. \tag{17}
$$

Corollary 7 *Under the assumptions of Theorem 1 and S1, the population values of the parameters in the optimal revision regression in Eq. (17) satisfy*

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

Eq. (17) can be re-written as a regression of the target variable on all of the forecasts, from h_1 to h_H, and the parameter restrictions given in Corollary 7 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 single-horizon MZ regression (13).

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. That is, we can undertake a MZ regression of the short-horizon forecast on a long-horizon forecast

$$\hat{Y}_{t|t-h_1} = \tilde{\alpha}_j + \tilde{\beta}_j \hat{Y}_{t|t-h_j} + v_{t|t-h_j} \quad \text{for all } h_j > h_1. \tag{18}$$

Similarly, we get a vector MZ test that uses the short-horizon forecasts as target variables:

$$\begin{bmatrix} \hat{Y}_{t+h_2|t+h_1} \\ \vdots \\ \hat{Y}_{t+h_H|t+h_{H-1}} \end{bmatrix} = \begin{bmatrix} \tilde{\alpha}_2 \\ \vdots \\ \tilde{\alpha}_H \end{bmatrix} + \begin{bmatrix} \tilde{\beta}_2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \tilde{\beta}_H \end{bmatrix} \begin{bmatrix} \hat{Y}_{t+h_2|t} \\ \vdots \\ \hat{Y}_{t+h_H|t} \end{bmatrix} + \begin{bmatrix} v_{t+h_2|t} \\ \vdots \\ v_{t+h_H|t} \end{bmatrix}. \tag{19}$$

Finally, we can estimate a version of the optimal revision regression:

$$\hat{Y}_{t|t-h_1} = \tilde{\alpha} + \tilde{\beta}_H \hat{Y}_{t|t-h_H} + \sum_{j=2}^{H-1} \tilde{\beta}_j \eta_{t|h_j,h_{j+1}} + v_t, \tag{20}$$

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

16

Corollary 8 *Under the assumptions of Theorem 1 and S1, (a) the population values of the para-meters in the MZ regression in Eq. (18) satisfy*

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

(b) the population values of the parameters in the vector MZ regression in Eq. (19) satisfy

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

(c) the population values of the parameters in the optimal revision regression in Eq. (20) satisfy

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

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 $E_{t-h_L}\left[\hat{Y}_{t|t-h_S}\right] = E_{t-h_L}[Y_t]$ 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.

3.5   Relation between regression and bounds tests

In this section we show that certain forms of sub-optimality will remain undetected by Mincer-Zarnowitz regressions, even in population, but can be detected using the bounds introduced in the previous section. Consider the following simple form for a sub-optimal forecast:

$$\tilde{Y}_{t|t-h} = \gamma_h + \lambda_h \hat{Y}_{t|t-h}^* + u_{t-h}, \text{ where } u_{t-h} \sim N\left(0, \sigma_{u,h}^2\right). \tag{21}$$

An optimal forecast would have $\left(\lambda_h, \gamma_h, \sigma_{u,h}^2\right) = (1, 0, 0)$. Certain combinations of these parameters will not lead to a rejection of the MZ null hypothesis, even when they deviate from $(1, 0, 0)$. Consider

the MZ regression:

$$Y_t = \alpha_h + \beta_h \tilde{Y}_{t|t-h} + \varepsilon_t$$

The population values of these parameters are

$$\beta_h = \frac{Cov\left[Y_t, \tilde{Y}_{t|t-h}\right]}{V\left[\tilde{Y}_{t|t-h}\right]} = \frac{\lambda_h V\left[\hat{Y}^*_{t|t-h}\right]}{\lambda_h^2 V\left[\hat{Y}^*_{t|t-h}\right] + \sigma^2_{u,h}},$$

so $\beta_h = 1$ implies $\sigma^2_{u,h} = \lambda_h (1 - \lambda_h) V\left[\hat{Y}^*_{t|t-h}\right]$. Moreover,

$$\alpha_h = E[Y_t] - \beta_h E\left[\tilde{Y}_{t|t-h}\right] = E[Y_t] - \beta_h (\gamma_h + \lambda_h E[Y_t]).$$

Hence, if $\beta_h = 1$, then $\alpha_h = 0$ provided that $\gamma_h = E[Y_t](1 - \lambda_h)$. Thus we can choose any $\lambda_h \in (0, 1)$ and find the parameters $\left(\gamma_h, \sigma^2_{u,h}\right)$ that lead to MZ parameters that satisfy the null of rationality. We now verify that such a parameter vector would violate one of the multi-horizon bounds. Consider the bound that the variance of the forecast should be decreasing in the horizon. In this example we have $V\left[\tilde{Y}_{t|t-h}\right] = \lambda_h^2 V\left[\hat{Y}^*_{t|t-h}\right] + \sigma^2_{u,h} = \lambda_h V\left[\hat{Y}^*_{t|t-h}\right]$, when $\sigma^2_{u,h} = \lambda_h (1 - \lambda_h) V\left[\hat{Y}^*_{t|t-h}\right]$. We know by rationality that $V\left[\hat{Y}^*_{t|t-h}\right]$ is decreasing in h, but since λ_h can take any value in $(0, 1)$, *and* this value can change across horizons, a violation may be found. Specifically, a violation of the decreasing forecast variance bound $V\left[\tilde{Y}_{t|t-h_S}\right] < V\left[\tilde{Y}_{t|t-h_L}\right]$ will be found if

$$\frac{\lambda_{h_S}}{\lambda_{h_L}} < \frac{V\left[\hat{Y}^*_{t|t-h_L}\right]}{V\left[\hat{Y}^*_{t|t-h_S}\right]}. \tag{22}$$

It is also possible to construct an example where a MZ test would detect a sub-optimal forecast but a bounds-based test would not. A simple example of this is any combination where $\left(\lambda_h, \gamma_h, \sigma^2_{u,h}\right) \neq \left(\lambda_h, E[Y_t](1 - \lambda_h), \lambda_h (1 - \lambda_h) V\left[\hat{Y}^*_{t|t-h}\right]\right)$, and where $V\left[\tilde{Y}_{t|t-h_S}\right] > V\left[\tilde{Y}_{t|t-h_L}\right]$. For example, $\left(\lambda_h, \gamma_h, \sigma^2_{u,h}\right) = (\lambda_h, 0, 0)$ for any $\lambda_{h_S} = \lambda_{h_L} = \lambda_h \in (0, 1)$. We summarize these examples in the following proposition:

Proposition 1 *The MZ regression test and variance bound tests do not subsume one another: Rejection of forecast optimality by one test need not imply rejection by the other.*

4   Extensions and Discussion

This section shows how our tests cover certain forms of non-stationary processes and heterogeneous forecast horizons and also shows that the results can be extended to a broader class of loss functions than squared error loss and can be used to detect model misspecification.

4.1   Stationarity and Tests of Forecast Optimality

The literature on forecast evaluation conventionally assumes that the underlying data generating process is covariance stationary. To see the role played by the covariance stationarity assumption, recall $\hat{Y}^*_{t+h|t-j} = \arg\min_{\hat{y}} E_{t-j}[(Y_{t+h} - \hat{y})^2]$. By optimality, we must have

$$E_t[(Y_{t+h} - \hat{Y}^*_{t+h|t-j})^2] \geq E_t[(Y_{t+h} - \hat{Y}^*_{t+h|t})^2] \quad \text{for } j \geq 1. \tag{23}$$

Then, by the law of iterated expectations,

$$E[(Y_{t+h} - \hat{Y}^*_{t+h|t-j})^2] \geq E[(Y_{t+h} - \hat{Y}^*_{t+h|t})^2] \quad \text{for } j \geq 1. \tag{24}$$

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, $t - j$, and does not require covariance stationarity. This uses a so-called "fixed event" set-up, where the target variable (Y_{t+h}) is kept fixed, and the horizon of the forecast is allowed to vary (from $t - j$ to t).

When the forecast errors are stationary, it follows from Eq. (24) that

$$E[(Y_{t+h+j} - \hat{Y}^*_{t+h+j|t})^2] \geq E[(Y_{t+h} - \hat{Y}^*_{t+h|t})^2] \quad \text{for } j \geq 1. \tag{25}$$

Equation (25) does not follow from Eq. (24) under non-stationarity. For example, suppose there is a deterministic reduction in the variance of Y between periods $\tau + h$ and $\tau + h + j$, such as:

$$Y_t = \begin{cases} \mu + \sigma \varepsilon_t & \text{for } t \leq \tau + h \\ \mu + \frac{\sigma}{2} \varepsilon_t & \text{for } t > \tau + h \end{cases}, \tag{26}$$

19

where ε_t is zero-mean white noise. This could be a stylized example of the "Great Moderation". Clearly Eq. (25) is now violated as $\hat{Y}^*_{\tau+h+j|\tau} = \hat{Y}^*_{\tau+h|\tau} = \mu$, and so

$$E[(Y_{\tau+h+j} - \hat{Y}^*_{\tau+h+j|\tau})^2] = \frac{\sigma^2}{4} < \sigma^2 = E[(Y_{\tau+h} - \hat{Y}^*_{\tau+h|\tau})^2] \quad \text{for } j \geq 1. \tag{27}$$

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 (unconditional expected squared) errors than, say, a three-year-ahead forecast (i.e. for GDP growth in 1985, after volatility has come down).

One way to deal with non-stationarities such as the break in the variance in Eq. (26) is to hold the date of the target variable fixed and to vary the forecast horizon as in fixed-event forecasts, see Clements (1997) and Nordhaus (1987). In this case the forecast optimality test is based on Eq. (24) rather than Eq. (25). To see how this works, notice that, by forecast optimality and the law of iterated expectations, for $h_L > h_S$

$$\begin{aligned} E_{t-h_S}[(Y_t - \hat{Y}^*_{t|t-h_L})^2] &\geq E_{t-h_S}[(Y_t - \hat{Y}^*_{t|t-h_S})^2], \quad \text{and} \\ E[(Y_t - \hat{Y}^*_{t|t-h_L})^2] &\geq E[(Y_t - \hat{Y}^*_{t|t-h_S})^2]. \end{aligned} \tag{28}$$

For the example with a break in the variance in Eq. (26), we have $\hat{Y}^*_{t|t-h_L} = \hat{Y}^*_{t|t-h_S} = \mu$, and

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

Using a fixed-event setup, we next show that the natural extensions of the inequality results established in Corollaries 1, 2, 3, and 4 also hold for a more general class of stochastic processes that do not require covariance stationarity but, rather, allows for unconditional heteroskedasticity such as in Eq. (26) and dependent, heterogeneously distributed data processes.

Proposition 2 *Define the following variables*

$$\overline{MSE}_T(h) \equiv \frac{1}{T} \sum_{t=1}^{T} MSE_t(h), \quad where \quad MSE_t(h) \equiv E\left[\left(Y_t - \hat{Y}^*_{t|t-h}\right)^2\right]$$

$$\overline{MSF}_T(h) \equiv \frac{1}{T} \sum_{t=1}^{T} MSF_t(h), \quad where \quad MSF_t(h) \equiv E\left[\hat{Y}^{*2}_{t|t-h}\right],$$

$$\overline{C}_T(h) \equiv \frac{1}{T} \sum_{t=1}^{T} C_t(h), \quad where \quad C_t(h) \equiv E\left[\hat{Y}^*_{t|t-h}Y_t\right]$$

$$\overline{MSFR}_T(h_S, h_L) \equiv \frac{1}{T} \sum_{t=1}^{T} MSFR_t(h_S, h_L), \quad where \quad MSFR_t(h_S, h_L) \equiv E\left[d^{*2}_{t|h_S,h_L}\right]$$

$$\overline{B}_T(h) \equiv \frac{1}{T} \sum_{t=1}^{T} B_t(h), \quad where \quad B_t(h_S, h_L) \equiv E\left[Y_t d^*_{t|h_S,h_L}\right]$$

then, under the assumptions of Theorem 1, the following bounds hold for any $h_S < h_M < h_L$:

$$(a) \qquad \overline{MSE}_T(h_S) \leq \overline{MSE}_T(h_L)$$

$$(b) \qquad \overline{MSF}_T(h_S) \geq \overline{MSF}_T(h_L)$$

$$(c) \qquad \overline{C}_T(h_S) \geq \overline{C}_T(h_L)$$

$$(d) \qquad \overline{MSFR}_T(h_S, h_M) \leq \overline{MSFR}_T(h_S, h_L)$$

$$(e) \qquad \overline{MSFR}_T(h_S, h_L) \leq 2\overline{B}_T(h_S, h_L)$$

Allowing for heterogeneity in the data does not affect the bounds obtained in Section 2 which were derived under the assumption of stationarity: rather than holding for the (unique) unconditional expectation, under data heterogeneity they hold for the unconditional expectation at each point in time, and for the average of these across the sample. The bounds for averages of unconditional moments presented in Proposition 2 can be tested by drawing on a central limit theorem for heterogeneous, serially dependent processes, see, e.g., Wooldridge and White (1988) and White (2001). The following proposition provides conditions under which these quantities can be estimated.

Proposition 3 *Define*

$$\delta_{jt}^{MSE} \equiv \left(Y_t - \hat{Y}^*_{t|t-h_j}\right)^2 - \left(Y_t - \hat{Y}^*_{t|t-h_{j-1}}\right)^2, \;\; for\; j = 2, ..., H$$

$$\delta_{jt}^{MSF} \equiv \hat{Y}^{*2}_{t|t-h_j} - \hat{Y}^{*2}_{t|t-h_{j-1}}, \;\; for\; j = 2, ..., H$$

$$\delta_{jt}^{C} \equiv Y_t\hat{Y}^*_{t|t-h_j} - Y_t\hat{Y}^*_{t|t-h_{j-1}}, \;\; for\; j = 2, ..., H$$

$$\delta_{jt}^{MSFR} \equiv \left(d^*_{t|h_1,h_j}\right)^2 - \left(d^*_{t|h_1,h_{j-1}}\right)^2, \;\; for\; j = 3, ..., H$$

$$\delta_{jt}^{B} \equiv Y_t d^*_{t|h_1,h_j} - Y_t d^*_{t|h_1,h_{j-1}}, \;\; for\; j = 3, ..., H$$

$$\boldsymbol{\delta}_t^k \equiv \left[\delta_{qt}^k, ..., \delta_{Ht}^k\right]', \;\; \hat{\boldsymbol{\Delta}}_T^k \equiv \frac{1}{T}\sum_{t=1}^T \boldsymbol{\delta}_t^{k\prime}, \;\; V_T^k \equiv V\left[\frac{1}{\sqrt{T}}\sum_{t=1}^T \boldsymbol{\delta}_t^{k\prime}\right],$$

where $k \in \{MSE, MSF, C, MSFR, B)$ and $q = 2$ for $k \in \{MSE, MSF, C\}$ and $q = 3$ for $k \in \{MSFR, B\}$. Assume: (i) $\boldsymbol{\delta}_t^k = \boldsymbol{\Delta}^k + \boldsymbol{\epsilon}_t^k$, for $t = 1, 2, ...,$ $\boldsymbol{\Delta} \in \mathbb{R}^{H-1}$; (ii) $\boldsymbol{\epsilon}_t^k$ is a uniform mixing sequence with ϕ of size $-r/2\,(r-1)$, $r \geq 2$ or a strong mixing sequence with α of size $-r/\,(r-2)$, $r > 2$; (iii) $E\left[\boldsymbol{\epsilon}_t^k\right] = 0$ for $t = 1, 2, ...T$; (iv) $E\left[\left|\epsilon_{it}^k\right|^r\right] < C < \infty$ for $i = 1, 2, ..., H-1$; (v) V_T^k is uniformly positive definite; (vi) There exists a $\hat{V}_T^k$ that is symmetric and positive definite such that $\hat{V}_T^k - V_T^k \to^p 0$. Then:

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

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.

4.2 Bounds for forecasts with heterogeneous horizons

Some economic data sets contain forecasts that have a wide variety of horizons, which the researcher may prefer to aggregate into a smaller set of forecasts. For example, the Greenbook forecasts we study in our empirical application are recorded at irregular times within a given quarter, so that the forecast labeled as a one-quarter horizon forecast, for example, may actually have a horizon of one, two or three months. Given limited time series observations it may not be desirable to attempt

to study all possible horizons, ranging from zero to 15 months. Instead, we may wish to aggregate these into forecasts of $h_S \in \{1, 2, 3\}$, $h_L \in \{4, 5, 6\}$, etc.

The proposition below shows that the inequality results established in the previous sections also apply to forecasts with heterogeneous horizons. The key to this proposition is that any "short" horizon forecast must have a corresponding "long" horizon forecast. With that satisfied, and ruling out correlation between the forecast error and whether a particular horizon length was chosen, we show that the bounds hold for heterogeneous forecast horizons. We state and prove the proposition below only for MSE; results for the other bounds follow using the same arguments.

Proposition 4 *Consider a data set of the form* $\left\{ \left(Y_t, \hat{Y}^*_{t|t-h_t}, \hat{Y}^*_{t|t-h_t-k_t} \right)' \right\}_{t=1}^T$, *where* $k_t > 0 \; \forall \; t$. *Let the assumptions of Theorem 1 and S1 hold. (a) If* (h_t, k_t) *are realizations from some stationary random variable and* $e^*_{t|t-h_j}$ *and* $e^*_{t|t-h_j-k_i}$ *are independent of* $\mathbf{1}\{h_t = h_j\}$ *and* $\mathbf{1}\{k_t = k_i\}$, *then:*

$$MSE_S \equiv E\left[\left(Y_t - \hat{Y}^*_{t|t-h_t} \right)^2 \right] \leq E\left[\left(Y_t - \hat{Y}^*_{t|t-h_t-k_t} \right)^2 \right] \equiv MSE_L$$

(b) If $\{h_t, k_t\}$ *is a sequence of pre-determined values, then:*

$$\overline{MSE}_{S,T} \equiv \frac{1}{T} \sum_{t=1}^T E\left[\left(Y_t - \hat{Y}^*_{t|t-h_t} \right)^2 \right] \leq \frac{1}{T} \sum_{t=1}^T E\left[\left(Y_t - \hat{Y}^*_{t|t-h_t-k_t} \right)^2 \right] \equiv \overline{MSE}_{L,T}$$

The only non-standard assumption here is in part (a), where we assume that process determining the short and long forecast horizons is independent of the forecast errors at those horizons. This rules out choosing particular (h_t, k_t) combinations after having inspected their resulting forecast errors, which could of course overturn the bounds. Notice that heterogeneity of the short and long forecast horizon lengths in part (b) induces heterogeneity in the mean squared errors, even when the data generating process is stationary. The assumption of stationarity of the data generating process in both parts can be relaxed using similar arguments as in Proposition 2.

4.3 Bounds under Bregman loss

The bounds in our paper are predicated upon the optimal forecast being the conditional mean. It is well-known that under squared error loss the optimal forecast is the conditional mean, but this

result in fact holds for a broader class of loss functions, known as "Bregman" loss, which take the form:

$$L\left(Y_{t+h}, \hat{Y}_{t+h|t}\right) = \phi\left(Y_{t+h}\right) - \phi\left(\hat{Y}_{t+h|t}\right) - \phi'\left(\hat{Y}_{t+h|t}\right)\left(Y_{t+h} - \hat{Y}_{t+h|t}\right), \qquad (29)$$

where ϕ is a strictly convex function, so $\phi'' > 0$.

It follows from the first order condition for an optimal forecast that any loss function in this class yields the conditional mean as the optimal forecast

$$
\begin{aligned}
0 &= E\left[\frac{\partial}{\partial \hat{y}} L\left(Y_{t+h}, \hat{Y}^*_{t+h|t}\right)|\mathcal{F}_t\right] \\
&= E\left[-\phi'\left(\hat{Y}^*_{t+h|t}\right) - \phi''\left(\hat{Y}^*_{t+h|t}\right)\left(Y_{t+h} - \hat{Y}^*_{t+h|t}\right) + \phi'\left(\hat{Y}^*_{t+h|t}\right)|\mathcal{F}_t\right] \\
&= -\phi''\left(\hat{Y}^*_{t+h|t}\right)\left(E\left[Y_{t+h}|\mathcal{F}_t\right] - \hat{Y}^*_{t+h|t}\right),
\end{aligned}
$$

so $\hat{Y}^*_{t+h|t} = E\left[Y_{t+h}|\mathcal{F}_t\right]$ since $\phi'' > 0$. This class was proposed by Bregman (1967), re-derived in Patton (2010) for strictly positive variables, and is discussed in Gneiting (2010). Figure 1 illustrates some loss functions that are members of this class. These can be asymmetric, convex or concave, yet all imply that the optimal forecast is the conditional mean, and thus all of the results in this paper apply to loss functions in this class.

4.4   Multi-horizon bounds and model misspecification

If a forecaster uses an internally-consistent but misspecified model to predict some target variable, will any of the tests presented above be able to detect it? We study this problem in two cases: one where the multi-step forecasts are obtained from a suite of horizon-specific models ("direct" multi-step forecasts), and the other where forecasts for all horizons are obtained from a single model (and multi-step forecasts are obtained by "iterating" on the one-step model). In both cases we show that model misspecification can indeed be detected using the multi-horizon bounds presented above.

4.4.1   Direct multi-step forecasts

If the forecaster is using different models for different forecast horizons it is perhaps not surprising that the resulting forecasts may violate one or more of the bounds presented in the previous section.

To illustrate this, consider a target variable that evolves according to a stationary AR(2) process,

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t, \quad \varepsilon_t \sim iid\ N\left(0, \sigma^2\right), \tag{30}$$

but the forecaster uses a direct projection of Y_t onto Y_{t-h} to obtain an h-step forecast:

$$Y_t = \rho_h Y_{t-h} + v_t, \quad \text{for } h = 1, 2, ... \tag{31}$$

Note that by the properties of an AR(2) we have:

$$\rho_1 = \frac{\phi_1}{1 - \phi_2}, \quad \rho_2 = \frac{\phi_1^2 - \phi_2^2 + \phi_2}{1 - \phi_2}. \tag{32}$$

For many combinations of (ϕ_1, ϕ_2) we obtain $|\rho_2| > |\rho_1|$, e.g., for $(\phi_1, \phi_2) = (0.1, 0.8)$ we find $\rho_1 = 0.5$ and $\rho_2 = 0.85$. This directly leads to a violation of the bound in Corollary 2, that the variance of the forecast should be weakly decreasing in the horizon. Further, it is simple to show that it also violates the MSE bound in Corollary 1:

$$MSE_1 \equiv E\left[\left(Y_t - \hat{Y}_{t|t-1}\right)^2\right] = \sigma_y^2\left(1 - \rho_1^2\right) = 0.75\sigma_y^2, \tag{33}$$

$$MSE_2 \equiv E\left[\left(Y_t - \hat{Y}_{t|t-2}\right)^2\right] = \sigma_y^2\left(1 - \rho_2^2\right) = 0.28\sigma_y^2.$$

In a situation such as this, the forecaster should recognize that the two-step forecasting model is better than the one-step forecasting model, and so simply use the two-step forecast again for the one-step forecast. (Or better yet, improve the forecasting models being used.)

4.4.2 Iterated multi-step forecasts

If a forecaster uses the same, misspecified, model to generate forecasts for all horizons it may seem unlikely that the resulting term structure of forecasts will violate one or more of the bounds presented above. We present here one simple example where this turns out to be true. Consider again a target variable that evolves according to a stationary AR(2) process as in Eq. (30), but the forecaster uses an AR(1) model:

$$Y_t = \rho_1 Y_{t-1} + v_t \tag{34}$$

$$\text{so } \hat{Y}_{t|t-h} = \rho_1^h Y_{t-h}, \quad \text{for } h = 1, 2, ...$$

where $\rho_1 = \phi_1 / (1 - \phi_2)$ is the population value of the AR(1) parameter when the DGP is an AR(2). (It is possible to show that a simple Mincer-Zarnowitz test, discussed in the next section, will *not* detect the use of a misspecified model, as the population parameters of the MZ regression in this case can be shown to satisfy $(\alpha, \beta) = (0, 1)$. However, a simple extension of the MZ regression that includes a lagged forecast error would be able to detect this model misspecification.) We now verify that this model misspecification may be detected using the bounds on MSE:

$$
\begin{aligned}
MSE_1 &\equiv E\left[\left(Y_t - \hat{Y}_{t|t-1}\right)^2\right] = \sigma_y^2 \left(1 - \rho_1^2\right), &(35)\\
MSE_2 &\equiv E\left[\left(Y_t - \hat{Y}_{t|t-2}\right)^2\right] = \sigma_y^2 \left(1 - \rho_1^4 + 2\rho_1^2 \left(\rho_1^2 - 1\right) \phi_2\right).
\end{aligned}
$$

Intuitively, if we are to find an AR(2) such that the one-step MSE from a misspecified AR(1) model is greater than that for a two-step forecast from the AR(1) model, it is likely a case where the true AR(1) coefficient is small relative to the AR(2) coefficient. Consider again the case that $(\phi_1, \phi_2) = (0.1, 0.8)$. The one- and two-step MSEs from the AR(1) model are then

$$
MSE_1 = 0.75\sigma_y^2, \quad MSE_2 = 0.64\sigma_y^2.
$$

Thus the MSE bound is violated. It is also possible to show that a test based only on the forecasts can detect model misspecification, despite the fact that the model is used in an internally consistent fashion across the forecast horizons. Consider the forecast revision, $d_{t|h_S,h_L} \equiv \hat{Y}_{t|t-h_S} - \hat{Y}_{t|t-h_L}$, and the result established in Corollary 1 that the mean squared forecast revision should be weakly increasing in the forecast horizon. Using the same AR(2) example we obtain:

$$
\begin{aligned}
MSFR_{1,2} &\equiv E\left[d_{t|1,2}^2\right] = \rho_1^2 \left(1 - \rho_1^2\right) \sigma_y^2 = 0.19\sigma_y^2,\\
MSFR_{1,3} &\equiv E\left[d_{t|1,3}^2\right] = \rho_1^2 \left(1 + \rho_1^4 - 2\rho_1^2\rho_2\right) \sigma_y^2 = 0.16\sigma_y^2.
\end{aligned}
$$

Hence the MSFR is not increasing from horizon 2 to 3, in violation of forecast optimality. Tests based on the MSE or MSFR bounds would detect, at least asymptotically, the use of a misspecified forecasting model, even though the model is being used consistently across horizons.

These simple examples illustrate that our variance bounds may be used to identify suboptimal forecasting models, even when being used consistently, and thus may help to spur improvements of misspecified forecasting models.

5   Monte Carlo Simulations

There is little existing evidence on the finite sample performance of forecast rationality tests, particularly when multiple forecast horizons are simultaneously involved. Moreover, our proposed set of rationality tests take the form of bounds on second moments of the data and can be implemented using the Wolak (1989) test of inequality constraints, the performance of which in time series applications such as ours is not well known. 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 scenarios. We first 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:

$$Y_t = \mu_y + \phi \left(Y_{t-1} - \mu_y \right) + \varepsilon_t, \ \ \varepsilon_t \sim iid \ N \left(0, \sigma_\varepsilon^2 \right) \ \text{for } t = 1, ..., T = 100. \tag{36}$$

The parameters are calibrated to quarterly US CPI inflation data: $\phi = 0.5$, $\sigma_y^2 = 0.5$, $\mu_y = 0.75$. Optimal forecasts for this process are given by $\hat{Y}^*_{t|t-h} = E_{t-h}[Y_t] = \mu_y + \phi^h \left(Y_{t-h} - \mu_y \right)$. We consider all horizons between $h = 1$ and $h = H$, and set $H \in \{ 4, 8 \}$.

5.1.1   Measurement error

The performance of rationality 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

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

Three values are considered for the magnitude of the measurement error, σ_ψ: (i) *zero*, $\sigma_\psi = 0$ (as for CPI); (ii) *medium*, $\sigma_\psi = \sqrt{0.7}\sigma_y$ (similar to GDP growth first release data as reported by Faust, Rogers and Wright (2005)); and (iii) *high*, $\sigma_\psi = 2\sqrt{0.7}\sigma_y$, which is chosen as twice the medium value.

5.1.2   Sub-optimal forecasts

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

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

where $\sigma_{\xi,h} = \sqrt{0.7}\sigma_y$ 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 $2 \times \sqrt{0.7}\sigma_y$ for the longest forecast horizon ($H = 8$):

$$\sigma_{\xi,h} = \frac{2\left(h-1\right)}{7} \times \sqrt{0.7}\sigma_y, \ \ \text{for } h = 1,2,...,H \leq 8.$$

This scenario is designed to mimic the situation where estimation error, or other sources of noise, are greater at longer horizons.

5.2   Simulation 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. Conventionally, Bonferroni bounds tests are conservative and tend to be undersized. Here, the individual MZ regression tests are severely oversized, and

the use of the Bonferroni bound partially mitigates this feature. 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 across the two scenarios. 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 very good power to detect this type of deviation from optimality, around 70-99%, particularly when the short-horizon forecast, $\hat{Y}_{t|t-1}$, which is not affected by noise, is used as the dependent variable. The covariance bound in Corollary 4 works so well because noise in the forecast increases $E[d^2_{t|h_S,h_L}]$ without affecting $E[Y_t d_{t|h_S,h_L}]$, thereby making it less likely that $E[2Y_t d_{t|h_S,h_L} - d^2_{t|h_S,h_L}] \geq 0$ holds. The univariate optimal revision regression in Eq. (17) also has excellent power properties, notably 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 Eq. (7). Thus, as expected, Panel B of Table 2 shows that this test has very good power under this scenario: 45% in the case with four forecast horizons, rising to 100% in the case with eight forecast horizons. The MSE and MSFR bounds have essentially zero power for this type of deviation from forecast optimality. The covariance bound based on the predicted variable has power around 15% when $H = 4$, which increases to a power of around 90% when $H = 8$. The covariance bound with the actual value replaced by the short-run forecast in Eq. (12), performs best among all tests, with power of 72% when $H = 4$ and power of 100% when $H = 8$. This is substantially higher than the power of the univariate optimal revision regression test in Eq. (17) which has power around 9-11% when conducted on the actual values and power of 53-66% when

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

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 below the nominal size, as expected for a Bonferroni-based test. However the power of the Bonferroni tests is high, and is comparable to the best of the individual tests. This suggests that it is possible and useful to combine the results of the various bounds-based tests via a simple Bonferroni test.

It is also possible to combine these tests into a single omnibus test by stacking the various inequalities into a single large vector and testing whether the weak inequality holds for all elements of this vector. We leave this approach aside for two reasons: The first relates to concerns about the finite-sample properties of a test with such a large number of inequalities relative to the number of available time series observations. The second relates to the interpretability of the omnibus test: by running each of the bounds tests separately we can gain valuable information into the sources of forecast sub-optimality, if present. An omnibus bounds test would, at most, allow us to state that a given sequence of forecasts is not optimal; it would not provide information on the direction of the sub-optimality.

We also used the bootstrap approaches of White (2000) and Hansen (2005) to implement the tests of forecast rationality based on multi-horizon bounds in our simulations and found that the finite-sample size and power from those approaches are very similar to those presented in Tables 1 and 2, and so we do not discuss them separately here.

In conclusion, 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 uni-variate optimal revision regression, particularly when the test uses the short-run forecast as the dependent variable. This test has good size and power properties and performs well across both

deviations from forecast efficiency. Across all tests, the covariance bound and the univariate optimal revision regression tests are the best individual tests. Our study also finds that Bonferroni bounds that combine the tests have good size and power properties.

6   Empirical Application

As an empirical illustration of the forecast optimality tests, we next evaluate the Federal Reserve "Greenbook" forecasts of quarter-over-quarter rates of change in GDP, the GDP deflator and CPI. Data are from Faust and Wright (2009), who extracted the Greenbook forecasts and actual values from real-time Fed publications. We use quarterly observations of the target variable over the period from 1981Q2 to 2000Q4, and the three series are plotted in Figure 2. The forecast series begin with the current quarter and run up to five quarters ahead in time, i.e., $h = 0, 1, 2, 3, 4, 5$. All series are reported in in annualized percentage points. If more than one Greenbook forecast is available within a given quarter, we use the earlier forecast. A few quarters have no forecasts at all, leaving a total of 89 periods with at least one forecast available. If we start the sample when a full set of six forecasts is available, and end it when the last full set of six forecasts is available, we are left with 79 observations. This latter approach is particularly useful for comparing the impact of the early part of our sample period, when inflation volatility was high.

The results of our tests of forecast rationality are reported in Table 3. Panel A shows the results for the sample that uses 79 observations, and represents our main empirical results. For GDP growth we observe a strong rejection of internal consistency via the univariate optimal revision regression using the short-run forecast as the target variable, Eq. (20), and a milder violation of the increasing mean-squared forecast revision test in Eq. (3). For the GDP deflator, several tests reject forecast optimality. In particular, the tests for decreasing covariance between the forecast and the actual, the covariance bound on forecast revisions, a decreasing mean squared forecast, and the univariate optimal revision regression all lead to rejections. Finally, for the CPI inflation rate we find a violation of the covariance bound, Eq. (12), and a rejection through the univariate optimal

revision regression. For all three variables, the Bonferroni combination test rejects multi-horizon forecast optimality at the 5% level.

Faust and Wright (2008) point out that forecasts from central banks are often based on an assumed path for the policy interest rate, and forecasts constructed using this assumption may differ from the central bank's best forecast of the target variable. In particular, they note that ignoring this conditioning can lead standard tests to over- or under-reject the null of forecast rationality. We leave the extension of our bounds-based tests to conditional forecasts for future work.

The sources of some of the rejections of forecast optimality are illustrated in Figures 3 and 4. For each of the series, Figure 3 plots the mean squared errors and variance of the forecasts. 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. In fact, the variance of the forecast 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 4 plots mean squared forecast revisions and the covariance between the forecast and the actual 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 rejection of forecast optimality for this case. In the right panel we see that while the covariance between the forecast and the actual is decreasing in the horizon for GDP growth and CPI, for the GDP deflator it is mostly flat, a contradiction of forecast rationality.

The Monte Carlo simulations are closely in line with our empirical findings. Rejections of forecast optimality come mostly from the covariance bound in Eq. (12) and the univariate optimal

revision regressions in Eq. (17) and Eq. (20). Moreover, for GDP growth, a series with greater measurement errors and data revisions, rejections tend to be stronger when only the forecasts are used.

In Panel B of Table 3 we present the results using the full sample of 89 observations, which gives us more short-horizon forecasts from the beginning of the sample period and more long-horizon forecasts from the end of the sample period. Changing the sample period does not greatly affect the results for GDP growth, although the rejection of rationality arising from the bound on mean-squared forecast revisions goes from being borderline significant to being borderline insignificant. The strong rejection of internal consistency via the univariate optimal revision regression using the short-run forecast as the target variable remains, as does the rejection using the Bonferroni bound to combine all tests. The results for the GDP deflator forecasts change, with three out of the five rejections using bounds tests vanishing, while the simple MZ test on the shortest horizon goes from not significant to strongly significant. The results for CPI inflation forecasts also change, with the bound on mean squared errors being significantly violated in this different sample period. Overall, this change in sample period does change the results of some of the individual tests, but the broader conclusions remain: for all three series, we find significant evidence against forecast rationality.

7   Conclusion

This paper proposes several new tests of forecast optimality that exploit information from multi-horizon forecasts. Our new tests are based on 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.

Our bounds hold whenever the optimal forecast is the conditional mean of the target variable. Beyond this case, 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 Granger (1999). The estimation of loss function parameters via moment (equality) conditions, as in Elliott, et al. (2005), may be combined with the multi-horizon inequality conditions presented above using recent work by Moon and Schorfheide (2009). Establishing results on multi-horizon forecasts when the loss function is unknown, as in Patton and Timmermann (2007b), may also be possible. We leave these extensions for future work.

8   Appendix A: Proofs

Proof of Corollary 1.   (a) By the optimality of $\hat{Y}^*_{t|t-h_S}$, and since $\hat{Y}^*_{t|t-h_L} \in \mathcal{F}_{t-h_S}$ for any $h_S < h_L$, 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 $E\left[e^{*2}_{t|t-h_S}\right] \leq E\left[e^{*2}_{t|t-h_L}\right]$ by the law of iterated expectations (LIE).   (b) Let $d^*_{t|h_S,h_L} \equiv \hat{Y}^*_{t|t-h_S} - \hat{Y}^*_{t|t-h_L} =$

$\left(\hat{Y}^*_{t|t-h_S} - \hat{Y}^*_{t|t-h_M} \right) + \left(\hat{Y}^*_{t|t-h_M} - \hat{Y}^*_{t|t-h_L} \right) = d^*_{t|h_S,h_M} + d^*_{t|h_M,h_L}$. Under the assumption that $h_S < h_M < h_L$ note that $E_{t-h_M}\left[d^*_{t|h_S,h_M} \right] = E_{t-h_M}\left[\hat{Y}^*_{t|t-h_S} - \hat{Y}^*_{t|t-h_M} \right] = 0$ by the LIE. Thus $Cov\left[d^*_{t|h_S,h_M}, d^*_{t|h_M,h_L} \right] = 0$ and so $V\left[d^*_{t|h_S,h_L} \right] = V\left[d^*_{t|h_S,h_M} \right] + V\left[d^*_{t|h_M,h_L} \right] \geq V\left[d^*_{t|h_S,h_M} \right]$. Further, since $E_{t-h}\left[d^*_{t|h,k} \right] = 0$ for any $h < k$ we then have $E\left[d^{*2}_{t|h_S,h_L} \right] \geq E\left[d^{*2}_{t|h_S,h_M} \right]$. $\blacksquare$

Proof of Corollary 2. Forecast optimality under MSE loss implies $\hat{Y}^*_{t|t-h} = E_{t-h}\left[Y_t \right]$. Thus $E_{t-h}\left[e^*_{t|t-h} \right] \equiv E_{t-h}\left[Y_t - \hat{Y}^*_{t|t-h} \right] = 0$, so $E\left[e^*_{t|t-h} \right] = 0$ and $Cov\left[\hat{Y}^*_{t|t-h}, e^*_{t|t-h} \right] = 0$, and $V\left[Y_t \right] = V\left[\hat{Y}^*_{t|t-h} \right] + E\left[e^{*2}_{t|t-h} \right]$, or $V\left[\hat{Y}^*_{t|t-h} \right] = V\left[Y_t \right] - E\left[e^{*2}_{t|t-h} \right]$. Corollary 1 showed that $E\left[e^{*2}_{t|t-h} \right]$ is weakly increasing in h, which implies that $V\left[\hat{Y}^*_{t|t-h} \right]$ must be weakly decreasing in h. $\blacksquare$

Proof of Corollary 3. As used in the above proofs, forecast optimality implies $Cov\left[\hat{Y}^*_{t|t-h}, e^*_{t|t-h} \right] = 0$ and thus $Cov\left[\hat{Y}^*_{t|t-h}, Y_t \right] = Cov\left[\hat{Y}^*_{t|t-h}, \hat{Y}^*_{t|t-h} + e^*_{t|t-h} \right] = V\left[\hat{Y}^*_{t|t-h} \right]$. Corollary 2 showed that $V\left[\hat{Y}^*_{t|t-h} \right]$ is weakly decreasing in h, and thus we have that $Cov\left[\hat{Y}^*_{t|t-h}, Y_t \right]$ is also weakly decreasing in h. For part (b), $Cov\left[\hat{Y}^*_{t|t-h_L}, \hat{Y}^*_{t|t-h_S} \right] = Cov\left[\hat{Y}^*_{t|t-h_L}, \hat{Y}^*_{t|t-h_L} + d^*_{t|h_S,h_L} \right] = V\left[\hat{Y}^*_{t|t-h_L} \right]$, since $Cov\left[\hat{Y}^*_{t|t-h_L}, d^*_{t|h_S,h_L} \right] = 0$. Similarly $Cov\left[\hat{Y}^*_{t|t-h_M}, \hat{Y}^*_{t|t-h_S} \right] = V\left[\hat{Y}^*_{t|t-h_M} \right]$, and the bound on covariances follows from the bound on $V\left[\hat{Y}^*_{t|t-h} \right]$ established in Corollary 2. $\blacksquare$

Proof of Corollary 4. For any $h_S < h_L$, Corollary 1 showed $V\left[Y_t - \hat{Y}^*_{t|t-h_L} \right] \geq V\left[Y_t - \hat{Y}^*_{t|t-h_S} \right]$, so $V\left[Y_t \right] + V\left[\hat{Y}^*_{t|t-h_L} \right] - 2Cov\left[Y_t, \hat{Y}^*_{t|t-h_L} \right] \geq V\left[Y_t \right] + V\left[\hat{Y}^*_{t|t-h_S} \right] - 2Cov\left[Y_t, \hat{Y}^*_{t|t-h_S} \right]$ and

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

Thus $V\left[d^*_{t|h_S,h_L} \right] \leq 2Cov\left[Y_t, d^*_{t|h_S,h_L} \right]$.

For the second part, $V\left[d^*_{t|h_M,h_L} \right] \leq 2Cov\left[Y_t, d^*_{t|h_M,h_L} \right] = 2Cov\left[\hat{Y}^*_{t|t-h_S} + e^*_{t|t-h_S}, d^*_{t|h_M,h_L} \right] = 2Cov\left[\hat{Y}^*_{t|t-h_S}, d^*_{t|h_M,h_L} \right]$ since $Cov\left[e^*_{t|t-h_S}, d^*_{t|h_M,h_L} \right] = 0$. $\blacksquare$

Proof of Corollary 5. (a) Note that $Cov\left[e^*_{t|t-h}, Y_t \right] = Cov\left[e^*_{t|t-h}, \hat{Y}^*_{t|t-h} + e^*_{t|t-h} \right] = V\left[e^*_{t|t-h} \right]$, since $Cov\left[e^*_{t|t-h}, \hat{Y}^*_{t|t-h} \right] = 0$ by forecast optimality. Corollary 1 established that MSE

is increasing in the forecast horizon, and thus we have that $Cov\left[e^*_{t|t-h}, Y_t\right]$ is increasing in the forecast horizon. (b) $Cov\left[d^*_{t|h_S,h_L}, \hat{Y}^*_{t|t-h_S}\right] = Cov\left[d^*_{t|h_S,h_L}, \hat{Y}^*_{t|t-h_L} + d^*_{t|h_S,h_L}\right] = V\left[d^*_{t|h_S,h_L}\right]$, since $Cov\left[d^*_{t|h_S,h_L}, \hat{Y}^*_{t|t-h_L}\right] = 0$. Corollary 1 established that the mean squared forecast error is increasing in the long forecast horizon, and thus we have that $Cov\left[d^*_{t|h_S,h_L}, \hat{Y}^*_{t|t-h_S}\right]$ is increasing in the long forecast horizon. (c) $Cov\left[e^*_{t|t-h_L}, d^*_{t|h_S,h_L}\right] \equiv Cov\left[Y_t - \hat{Y}^*_{t|t-h_L}, \hat{Y}^*_{t|t-h_S} - \hat{Y}^*_{t|t-h_L}\right] = Cov\left[Y_t, \hat{Y}^*_{t|t-h_S} - \hat{Y}^*_{t|t-h_L}\right]$, since $Cov\left[\hat{Y}^*_{t|t-h_L}, d^*_{t|h_S,h_L}\right] = 0$. Corollary 3 showed that $Cov\left[Y_t, \hat{Y}^*_{t|t-h}\right]$ is decreasing in the forecast horizon, thus, keeping h_S fixed, we find that $Cov\left[Y_t, \hat{Y}^*_{t|t-h_S} - \hat{Y}^*_{t|t-h_L}\right] = Cov\left[Y_t, \hat{Y}^*_{t|t-h_S}\right] - Cov\left[Y_t, \hat{Y}^*_{t|t-h_L}\right]$ is increasing in the long forecast horizon, as claimed. ∎

Proof of Corollary 6. The population value of β_h is $Cov\left[\hat{Y}_{t|t-h}, Y_t\right]/V\left[\hat{Y}_{t|t-h}\right]$, which under optimality equals $\beta_h = Cov\left[\hat{Y}^*_{t|t-h}, Y_t\right]/V\left[\hat{Y}^*_{t|t-h}\right] = Cov\left[\hat{Y}^*_{t|t-h}, \hat{Y}^*_{t|t-h} + e^*_{t|t-h}\right]/V\left[\hat{Y}^*_{t|t-h}\right] = V\left[\hat{Y}^*_{t|t-h}\right]/V\left[\hat{Y}^*_{t|t-h}\right] = 1$. The population value of α_h under optimality equals $\alpha_h = E\left[Y_t\right] - \beta_h E\left[\hat{Y}^*_{t|t-h}\right] = E\left[Y_t\right] - E\left[\hat{Y}^*_{t|t-h}\right] = 0$ by the LIE since $\hat{Y}^*_{t|t-h} = E_{t-h}\left[Y_t\right]$. ∎

Proof of Corollary 7. Let the parameters in the regression

$$Y_t = \alpha + \beta_H \hat{Y}_{t|t-h_H} + \sum_{j=1}^{H-1} \beta_j d_{t|h_j,h_{j+1}} + u_t,$$

be denoted $\theta = \left[\alpha, \beta_H, \beta_1, ..., \beta_{H-1}\right]'$. The result follows from the fact that the probability limit of the OLS estimator of θ is:

$$
\begin{pmatrix}
\bar{Y}^2_{h_H}/\sigma^2_{\hat{Y}_{h_H}} + 1 & -\bar{Y}_{h_H}/\sigma^2_{\hat{Y}_{h_H}} & 0 & \cdots & 0 \\
-\bar{Y}_{h_H}/\sigma^2_{\hat{Y}_{h_H}} & \sigma^{-2}_{\hat{Y}_{h_H}} & 0 & \cdots & 0 \\
0 & 0 & \sigma^{-2}_{\eta_{h_1,h_2}} & 0 & 0 \\
\vdots & \vdots & 0 & \ddots & \vdots \\
& & & & 0 \\
0 & 0 & 0 & \cdots & \sigma^{-2}_{\eta_{h_H-2,h_H-1}}
\end{pmatrix}
\begin{pmatrix}
\bar{Y}_{h_H} \\
\bar{Y}^2_{h_H} + \sigma^2_{\hat{Y}_{h_H}} \\
\sigma^2_{\eta_{h_1,h_2}} \\
0 \\
\vdots \\
\sigma^2_{\eta_{h_H-2,h_H-1}}
\end{pmatrix}
=
\begin{pmatrix}
0 \\
1 \\
\vdots \\
\\
1
\end{pmatrix},
$$

where $\bar{Y}_{h_H} = E[\hat{Y}_{t|t-h_H}]$, $\sigma^2_{\hat{Y}_{h_H}} = V\left[\hat{Y}_{t|t-h_H}\right]$ and we used properties of the partitioned inverse. ∎

Proof of Corollary 8. (a) Under optimality, $\tilde{\beta}_h = Cov\left[\hat{Y}^*_{t|t-h_1}, \hat{Y}^*_{t|t-h_j}\right]/V\left[\hat{Y}^*_{t|t-h_j}\right] = Cov\left[\hat{Y}^*_{t|t-h_j} + d^*_{t|h_1,h_j}, \hat{Y}^*_{t|t-h_j}\right]/V\left[\hat{Y}^*_{t|t-h_j}\right] = V\left[\hat{Y}^*_{t|t-h_j}\right]/V\left[\hat{Y}^*_{t|t-h_j}\right] = 1$, and $\tilde{\alpha}_h = E\left[\hat{Y}^*_{t|t-h_1}\right] -$

$\tilde{\beta}_h E\left[\hat{Y}^*_{t|t-h_j}\right] = E\left[Y_t\right] - E\left[Y_t\right] = 0.$ (b) Follows using the same steps as the proof of part

(b) of Corollary 7, noting that $\hat{Y}^*_{t|t-h_2} = E_{t-h_2}\left[Y_t\right] = E_{t-h_2}\left[\hat{Y}^*_{t|t-h_1}\right]$ by the LIE, and that

$E_{t-h_2}\left[\hat{Y}^*_{t|t-h_1} - \hat{Y}^*_{t|t-h_2}\right] \equiv E_{t-h_2}\left[d^*_{t|h_1,h_2}\right] = 0.$ ∎

Proof of Proposition 2. (a) By forecast optimality we have $E_{t-h_S}\left[\left(Y_t - \hat{Y}^*_{t|t-h_S}\right)^2\right] \leq$

$E_{t-h_S}\left[\left(Y_t - \hat{Y}^*_{t|t-h_L}\right)^2\right]$ for all $h_S < h_L$, 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. Thus $T^{-1}\sum_{t=1}^{T} E\left[\left(Y_t - \hat{Y}^*_{t|t-h_S}\right)^2\right] \leq T^{-1}\sum_{t=1}^{T} E\left[\left(Y_t - \hat{Y}^*_{t|t-h_L}\right)^2\right]$ as claimed.

(b) By forecast optimality we have $E_{t-h}\left[e^*_{t|t-h}\right] = 0$ so $E_{t-h}\left[e^*_{t|t-h}\hat{Y}^*_{t|t-h}\right] = 0 \Rightarrow E\left[e^*_{t|t-h}\hat{Y}^*_{t|t-h}\right] = $

0 by the LIE. This implies that $E\left[Y_t^2\right] = E\left[\left(\hat{Y}^*_{t|t-h} + e^*_{t|t-h}\right)^2\right] = E\left[\hat{Y}^{*2}_{t|t-h}\right] + E\left[e^{*2}_{t|t-h}\right] \equiv$

$MSF_t\left(h\right) + MSE_t\left(h\right)$, so $MSF_t\left(h\right) = E\left[Y_t^2\right] - MSE_t\left(h\right)$. We established in part (a) that

$MSE_t\left(h_S\right) \leq MSE_t\left(h_L\right)$ for all $h_S < h_L$ and for each t, and since $E\left[Y_t^2\right]$ is not a function of h,

this implies that $MSF_t\left(h_S\right) \geq MSF_t\left(h_L\right)$ for all $h_S < h_L$. Averaging over $t = 1, 2, .., T$ leads to

$\overline{MSF}_T\left(h_S\right) \geq \overline{MSF}_T\left(h_L\right)$.

(c) $E_{t-h}\left[Y_t\hat{Y}^*_{t|t-h}\right] = E_{t-h}\left[\left(\hat{Y}^*_{t|t-h} + e^*_{t|t-h}\right)\hat{Y}^*_{t|t-h}\right] = E_{t-h}\left[\hat{Y}^{*2}_{t|t-h}\right] \equiv MSF_t\left(h\right)$, since

$E_{t-h}\left[e^*_{t|t-h}\hat{Y}^*_{t|t-h}\right] = 0$ by optimality. We showed that $MSF_t\left(h\right)$ is decreasing in h in part (b),

thus $C_t\left(h\right) \equiv E_{t-h}\left[Y_t\hat{Y}^*_{t|t-h}\right]$ is also decreasing in h.

(d) By the fact that $\hat{Y}^*_{t|t-h} = E_{t-h}\left[Y_t\right]$ we have $E_{t-h_L}\left[\hat{Y}^*_{t|t-h_S} - \hat{Y}^*_{t|t-h_L}\right] \equiv E_{t-h_L}\left[d^*_{t|h_S,h_L}\right] =$

0 for all $h_S < h_L$. Thus $E_{t-h_M}\left[d^{*2}_{t|h_S,h_L}\right] = E_{t-h_M}\left[\left(d^*_{t|h_S,h_M} + d^*_{t|h_M,h_L}\right)^2\right] = E_{t-h_M}\left[d^{*2}_{t|h_S,h_M}\right] +$

$E_{t-h_M}\left[d^{*2}_{t|h_M,h_L}\right]$, and $E\left[d^{*2}_{t|h_S,h_L}\right] = E\left[d^{*2}_{t|h_S,h_M}\right] + E\left[d^{*2}_{t|h_M,h_L}\right]$ by the LIE. Thus $MSFR_t\left(h_S,h_L\right) \geq$

$MSFR_t\left(h_S,h_M\right)$ for all $h_S < h_M < h_L$ for each t. Averaging over $t = 1, 2, .., T$ leads to

$\overline{MSFR}_T\left(h_S,h_M\right) \leq \overline{MSFR}_T\left(h_S,h_L\right)$ as claimed.

(e) From (a) we have

$$E_{t-h_S}\left[\left(Y_t - \hat{Y}^*_{t|t-h_L}\right)^2\right] \geq E_{t-h_S}\left[\left(Y_t - \hat{Y}^*_{t|t-h_S}\right)^2\right] \forall h_S < h_L, \text{ which implies}$$

$$E_{t-h_S}\left[\hat{Y}^{*2}_{t|t-h_L}\right] - 2E_{t-h_S}\left[Y_t\hat{Y}^*_{t|t-h_L}\right] \geq E_{t-h_S}\left[\hat{Y}^{*2}_{t|t-h_S}\right] - 2E_{t-h_S}\left[Y_t\hat{Y}^*_{t|t-h_S}\right]$$

$$= E_{t-h_S}\left[\hat{Y}^{*2}_{t|t-h_L}\right] + E_{t-h_S}\left[d^{*2}_{t|h_S,h_L}\right] + 2E_{t-h_S}\left[\hat{Y}^*_{t|t-h_L}d^*_{t|h_S,h_L}\right]$$

$$-2E_{t-h_S}\left[Y_t\hat{Y}^*_{t|t-h_L}\right] - 2E_{t-h_S}\left[Y_t d^*_{t|h_S,h_L}\right]$$

37

So $E_{t-h_S}\left[d_{t|h_S,h_L}^{*2}\right] \leq 2E_{t-h_S}\left[Y_t d_{t|h_S,h_L}^*\right] \Rightarrow E\left[d_{t|h_S,h_L}^{*2}\right] \leq 2E\left[Y_t d_{t|h_S,h_L}^*\right]$ by the LIE. Averaging over $t = 1, 2, .., T$ leads to $\overline{MSFR}_T(h_S, h_L) \leq 2\overline{B}_T(h_S, h_L)$ as claimed. $\blacksquare$

Proof of Proposition 3. Follows from Exercise 5.21 of White (2001). $\blacksquare$

Proof of Proposition 4. (a) Let h_t, k_t be integer-valued random variables with support $\mathcal{H} \times \mathcal{K}$. Let $\Pr[h_t = h_j] = p_j$ and let $\Pr[k_t = k_i | h_t = h_j] = q_{ij}$. First, note that $e_{t|t-h_t}^* \equiv Y_t - \hat{Y}_{t|t-h_t}^* = \sum_{h_j \in \mathcal{H}}\left(Y_t - \hat{Y}_{t|t-h_j}^*\right) \cdot \mathbf{1}\{h_t = h_j\}$ and $e_{t|t-h_t}^{*2} = \sum_{h_j \in \mathcal{H}}\left(Y_t - \hat{Y}_{t|t-h_j}^*\right)^2 \cdot \mathbf{1}\{h_t = h_j\}$ since $\mathbf{1}\{h_t = h_j\}\mathbf{1}\{h_t = h_i\} = 0$ for $h_j \neq h_i$. Then the "short horizon" MSE equals:

$$MSE_S \equiv E\left[\left(Y_t - \hat{Y}_{t|t-h_t}^*\right)^2\right] = \sum_{h_j \in \mathcal{H}} E\left[\left(Y_t - \hat{Y}_{t|t-h_j}^*\right)^2 \cdot \mathbf{1}\{h_t = h_j\}\right] = \sum_{h_j \in \mathcal{H}} p_j E\left[\left(Y_t - \hat{Y}_{t|t-h_j}^*\right)^2\right]$$

while the "long horizon" MSE equals:

$$\begin{aligned}
MSE_L &\equiv E\left[\left(Y_t - \hat{Y}_{t|t-h_t-k_t}^*\right)^2\right] \\
&= \sum_{h_j \in \mathcal{H}}\sum_{k_i \in \mathcal{K}} E\left[\mathbf{1}\{k_t = k_i\}\mathbf{1}\{h_t = h_j\}\left(Y_t - \hat{Y}_{t|t-h_j-k_i}^*\right)^2\right] \\
&= \sum_{h_j \in \mathcal{H}}\sum_{k_i \in \mathcal{K}} E\left[\mathbf{1}\{k_t = k_i\}\mathbf{1}\{h_t = h_j\}\left(Y_t - \hat{Y}_{t|t-h_j-k_i}^*\right)^2 | h_t = h_j\right]\Pr[h_t = h_j] \\
\\
&= \sum_{h_j \in \mathcal{H}} p_j \sum_{k_i \in \mathcal{K}} E\left[\mathbf{1}\{k_t = k_i\}\left(Y_t - \hat{Y}_{t|t-h_j-k_i}^*\right)^2 | h_t = h_j\right] \\
&= \sum_{h_j \in \mathcal{H}} p_j \sum_{k_i \in \mathcal{K}} q_{ij} E\left[\left(Y_t - \hat{Y}_{t|t-h_j-k_i}^*\right)^2\right]
\end{aligned}$$

The difference between these is

$$\begin{aligned}
MSE_S - MSE_L &= \sum_{h_j \in \mathcal{H}} p_j \left(E\left[\left(Y_t - \hat{Y}_{t|t-h_j}^*\right)^2\right] - \sum_{k_i \in \mathcal{K}} q_{ij} E\left[\left(Y_t - \hat{Y}_{t|t-h_j-k_i}^*\right)^2\right]\right) \\
&= \sum_{h_j \in \mathcal{H}} p_j \left(\sum_{k_i \in \mathcal{K}} q_{ij}\left\{E\left[\left(Y_t - \hat{Y}_{t|t-h_j}^*\right)^2\right] - E\left[\left(Y_t - \hat{Y}_{t|t-h_j-k_i}^*\right)^2\right]\right\}\right) \\
&\leq 0 \text{ since } E\left[\left(Y_t - \hat{Y}_{t|t-h_j}^*\right)^2\right] \leq E\left[\left(Y_t - \hat{Y}_{t|t-h_j-k_i}^*\right)^2\right] \forall h_j, k_i.
\end{aligned}$$

(b) Let the short horizon lengths, h_t, and long horizon lengths, $h_t + k_t$, be given by some pre-determined sequence $\{h_t, k_t\}_{t=-\infty}^{\infty}$. Note that this introduces heterogeneity into the problem,

even when the underlying data generating process is stationary. Nevertheless, the bounds established in Section 2 continue to hold: By forecast optimality we have $E_{t-h_t} \left[\left(Y_t - \hat{Y}^*_{t|t-h_t} \right)^2 \right] \leq E_{t-h_t} \left[\left(Y_t - \hat{Y}^*_{t|t-h_t-k_t} \right)^2 \right]$ for each t, which implies $E \left[\left(Y_t - \hat{Y}^*_{t|t-h_t} \right)^2 \right] \leq E \left[\left(Y_t - \hat{Y}^*_{t|t-h_t-k_t} \right)^2 \right]$ by the LIE. Averaging over $t = 1, 2, .., T$ leads to a bound on the average MSE over the sample period $\overline{MSE}_T (h_S) \leq \overline{MSE}_T (h_L)$. Corresponding results can be obtain for the remaining bounds using the same arguments. ∎

9 Appendix B: Illustration of bounds for an AR(1) process

This Appendix illustrates the moment bounds for an AR(1) process. Let:

$$Y_t = \phi Y_{t-1} + \varepsilon_t, \quad |\phi| < 1, \tag{37}$$

where $\varepsilon_t \sim WN(0, \sigma^2_\varepsilon)$, so $\sigma^2_y = \sigma^2_\varepsilon / (1 - \phi^2)$. 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,

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

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

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

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

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

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 $d^*_{t|h_S, h_L} = \sum_{i=h_S}^{h_L-1} \phi^i \varepsilon_{t-i}$, and so

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

which is increasing in $h_L - h_S$, consistent with Corollary 1. Consistent with Corollary 4, the variance of the revision is bounded by twice the covariance of the actual value and the revision:

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

The implications of forecast rationality presented in Corollary 4 based on the predicted as opposed to realized value for Y for this AR(1) example are:

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

and

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

References

[1] Bregman, L.M., 1967, The Relaxation Method of Finding the Common Point of Convex Sets and its Application to the Solution of Problems in Convex Programming, USSR Computational Mathematics and Mathematical Physics, 7, 200-217.

[2] Capistran, C., 2007, Optimality Tests for Multi-Horizon Forecasts. Manuscript no 2007-14, Banco de Mexico.

[3] Clements, M.P., 1997, Evaluating the Rationality of Fixed-Event Forecasts. Journal of Forecasting 16, 225-239.

[4] Clements, M.P., and D.F. Hendry, 1998, Forecasting Economic Time Series. Cambridge. Cambridge University Press.

[5] Corradi, V., A. Fernandez, and N.R. Swanson, 2009, Information in the Revision Process of Real-Time Datasets, Journal of Business Economics and Statistics 27, 455-467.

[6] Croushore, D., 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, D. and T. 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, F.X., 2001, Elements of Forecasting. 2nd edition. Ohio: South-Western.

[10] Diebold, F.X. and G.D. Rudebusch, 1991, Forecasting Output with the Composite Leading Index: A Real-Time Analysis, Journal of the American Statistical Association, 86, 603-610.

[11] Elliot, G, I. Komunjer and A. Timmermann, 2005, Estimation and Testing of Forecast Rationality under Flexible Loss, Review of Economic Studies, 72(4), 1107-1125.

[12] Faust, J., J. Rogers and J. Wright, 2005, News and Noise in G-7 GDP Announcements. Journal of Money, Credit and Banking 37, 403-419.

[13] Faust, J. and J. Wright, 2008, Efficient Forecast Tests for Conditional Policy Forecasts, Journal of Econometrics, 146, 293-303.

[14] Faust, J. and J. Wright, 2009, Comparing Greenbook and Reduced Form Forecasts using a Large Realtime Dataset. Journal of Business and Economic Statistics 27, 468-479.

[15] Gneiting, T., 2010, Making and Evaluating Point Forecasts, working paper, Institut für Angewandte Mathematik, Universität Heidelberg.

[16] 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.

[17] Granger, C.W.J., 1969, Prediction with a Generalized Cost Function. OR 20, 199-207.

[18] Granger, C.W.J., 1999, Outline of Forecast Theory Using Generalized Cost Functions. Spanish Economic Review 1, 161-173.

[19] Hansen, P.R., 2005, A Test for Superior Predictive Ability. Journal of Business and Economic Statistics 23, 365-380.

[20] Isiklar, G., K. Lahiri and P. Loungani, 2006, How Quickly Do Forecasters Incorporate News? Evidence from Cross-country Surveys. Journal of Applied Econometrics 21, 703-725.

[21] Kudo, A., 1963, A Multivariate Analogue of the One-Sided Test, Biometrika, 50(3-4), 403-418.

[22] Marcellino, M., J.H. Stock and M.W. Watson, 2006, A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series.' Journal of Econometrics 135, 499–526.

[23] Mincer, J., and V. Zarnowitz, 1969, The Evaluation of Economic Forecasts, in J. Mincer (ed.) Economic Forecasts and Expectations, National Bureau of Economic Research, New York.

[24] Moon, H.R., and F. Schorfheide, 2009, Estimation with Over-Identifying Inequality Moment Conditions, Journal of Econometrics, 153, 136-154.

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

[26] Patton, A. J., 2010, Volatility Forecast Comparison Using Imperfect Volatility Proxies, Journal of Econometrics, forthcoming.

[27] Patton, A. J., and A. Timmermann, 2007a, Properties of Optimal Forecasts under Asymmetric Loss and Nonlinearity. Journal of Econometrics 140(2), 884-918.

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

[29] Patton, A. J., and A. Timmermann, 2008, Predictability of output growth and inflation: A multi-horizon survey approach. Forthcoming in the Journal of Business and Economic Statistics.

[30] Patton, A. J., and A. Timmermann, 2010a, Monotonicity in Asset Returns; New Tests with Applications to the Term Structure, the CAPM, and Portfolio Sorts, Journal of Financial Economics, 98, 605-625.

[31] Patton, A. J., and A. Timmermann, 2010b, Why do Forecasters Disagree? Lessons from the Term Structure of Cross-sectional Dispersion. Journal of Monetary Economics 57, 803-820.

[32] Pesaran, M.H. and M. Weale, 2006, Survey Expectations. Pages 715-776 in G. Elliott, C.W.J. Granger and A. Timmermann (eds.) Handbook of Economic Forecasting, North Holland: Amsterdam.

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

[34] Sun, H.-J., 1988, A General Reduction Method for n-Variate Normal Orthant Probability, Communications in Statistics, 17(11), 3913-3921.

[35] West, K.D., 1996, Asymptotic Inference about Predictive Ability. Econometrica 64, 1067-84.

[36] West, K.D., and M.W. McCracken, 1998, Regression Based Tests of Predictive Ability. International Economic Review 39, 817-840.

[37] White, H., 2000, A Reality Check for Data Snooping. Econometrica 68, 1097-1126.

[38] White, H., 2001, Asymptotic Theory for Econometricians, Second Edition, Academic Press, San Diego.

[39] Wolak, F.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.

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

[41] Wooldridge, J.M., and H. 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**

	H = 4			H = 8		
Meas. error variance:	*High*	*Med*	*Zero*	*High*	*Med*	*Zero*
Inc MSE	3.0	1.5	1.0	6.3	5.2	5.2
Dec COV	1.1	0.9	0.8	5.0	4.7	4.4
COV bound	1.8	1.4	1.2	0.0	0.0	0.0
Dec MSF	2.0	2.0	2.0	0.7	0.7	0.7
Inc MSFR	0.1	0.1	0.1	4.4	4.4	4.4
Dec COV, with proxy	1.2	1.2	1.2	6.0	6.0	6.0
COV bound, with proxy	3.8	3.8	3.8	0.0	0.0	0.0
MZ on short horizon	10.8	11.9	13.6	10.8	11.9	13.6
Univar opt. revision regr.	10.2	9.7	9.8	11.5	10.2	9.4
Univar opt. revision regr., with proxy	10.8	10.8	10.8	9.5	9.5	9.5
Univar MZ, Bonferroni	12.5	12.9	18.2	18.4	19.1	22.4
Univar MZ, Bonferroni, with proxy	17.8	17.8	17.8	20.8	20.8	20.8
Vector MZ	33.2	31.5	28.9	92.2	89.9	83.5
Vector MZ, with proxy	20.7	20.7	20.7	68.6	68.6	68.6
Bonf, using actuals	3.0	2.7	2.5	8.0	7.5	8.5
Bonf, using forecasts only	3.0	3.0	3.0	7.0	7.0	7.0
Bonf, all tests	3.7	3.2	2.3	8.1	7.2	6.1

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. Tests labeled "with proxy" 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

| | PANEL A: Equal noise across all forecast horizons | | | | | | PANEL B: Increasing noise across forecast horizons | | | | | |
| | H = 4 | | | H = 8 | | | H = 4 | | | H = 8 | | |
Meas. error variance:	High	Med	Zero	High	Med	Zero	High	Med	Zero	High	Med	Zero
Inc MSE	7.1	6.5	5.0	15.8	14.4	13.2	0.5	0.2	0.1	0.1	0.4	0.1
Dec COV	6.0	5.1	4.9	14.9	13.8	13.0	3.4	3.4	4.0	12.1	11.2	12.5
COV bound	72.4	78.0	82.5	73.5	78.9	84.0	13.3	14.6	16.0	89.1	91.6	96.0
Dec MSF	6.0	6.0	6.0	18.2	18.2	18.2	45.2	45.2	45.2	100.0	100.0	100.0
Inc MSFR	8.1	8.1	8.1	16.7	16.7	16.7	0.0	0.0	0.0	0.0	0.0	0.0
Dec COV, with proxy	8.4	8.4	8.4	15.5	15.5	15.5	5.0	5.0	5.0	13.9	13.9	13.9
COV bound, with proxy	98.5	98.5	98.5	99.2	99.2	99.2	72.7	72.7	72.7	100.0	100.0	100.0
MZ on short horizon	92.6	98.0	100.0	92.6	98.0	100.0	10.8	11.9	13.6	10.8	11.9	13.6
Opt. revision regr.	84.4	94.0	99.6	73.9	88.0	99.0	9.0	8.6	11.0	9.3	9.9	11.6
Opt. revision regr., proxy	100.0	100.0	100.0	100.0	100.0	100.0	66.1	66.1	66.1	53.9	53.9	53.9
Bonf, using actuals	68.4	83.8	97.7	67.3	79.3	95.4	12.7	12.2	12.8	99.4	99.4	99.5
Bonf, using forecasts only	100.0	100.0	100.0	100.0	100.0	100.0	63.9	63.9	63.9	100.0	100.0	100.0
Bonf, all tests	100.0	100.0	100.0	100.0	100.0	100.0	52.3	52.1	52.0	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 studied against sub-optimal forecasts obtained when forecasts are contaminated by the same level of noise across all horizons (Panel A) and when forecasts are contaminated by noise that increases in the horizon (Panel B). The simulations assume a sample of 100 observations and a nominal size of 10%. Tests labeled "with proxy" refer to cases where the one-period forecast is used in place of the predicted variable.

Table 3: Forecast rationality tests for Greenbook forecasts

Series:	Observations lined up in event time (T=79)			All available observations (T=89)		
	Growth	Deflator	CPI	Growth	Deflator	CPI
Inc MSE	0.591	0.966	0.639	0.423	0.922	0.000*
Dec COV	0.879	0.057*	0.991	0.838	0.670	0.896
COV bound	0.560	0.000*	0.009*	0.715	0.000*	0.036*
Dec MSF	0.916	0.026*	0.719	0.973	0.554	0.699
Inc MSFR	0.089*	0.938	0.620	0.123	0.938	0.340
Dec COV, with proxy	0.807	0.075*	0.772	0.811	0.375	0.632
COV bound, with proxy	0.206	0.010*	0.656	0.218	0.039*	0.671
MZ on short horizon	0.245	0.313	0.699	0.121	0.012*	0.037*
Opt. revision regr.	0.709	0.000*	0.001*	0.709	0.000*	0.001*
Opt. revision regr., with proxy	0.000*	0.009*	0.022*	0.000*	0.009*	0.022*
Bonf, using actuals	1.000	0.000*	0.005*	0.605	0.000*	0.000*
Bonf, using forecasts only	0.000*	0.047*	0.108	0.000*	0.047*	0.108
Bonf, all tests	**0.000***	**0.001***	**0.011***	**0.000***	**0.000***	**0.000***

Note: This table presents p-values from inequality- and regression tests of forecast rationality 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, 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 critical values based on a mixture of chi-squared variables. Tests labeled "with proxy" refer to cases where the shortest-horizon forecast forecast is used in place of the target variable in the test. P-values less than 0.10 are marked with an asterisk.
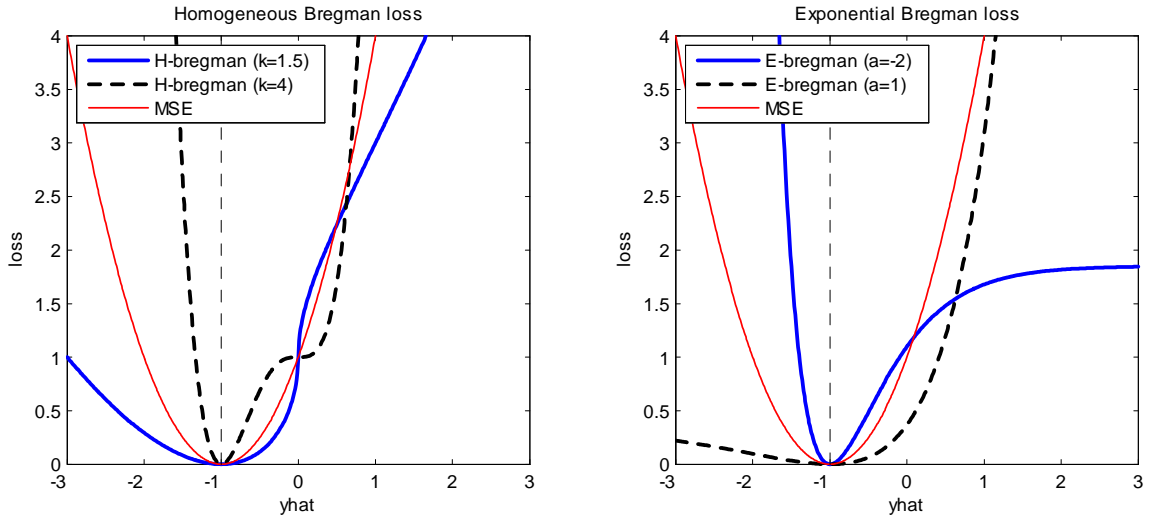
Figure 1: *Examples of Bregman loss functions for $\hat{y} \in [-3, 3]$ when $y = -1$. The left panel presents examples of "homogeneous Bregman" loss functions of the form $L(y, \hat{y}) = |y|^a - |\hat{y}|^a - \text{sgn}(\hat{y}) \, a \, |\hat{y}|^{a-1} (y - \hat{y})$, for $a > 1$. The right panel presents examples of "exponential Bregman" loss functions of the form $L(y, \hat{y}) = 2a^{-2} (\exp\{ay\} - \exp\{a\hat{y}\}) - 2a^{-1} \exp\{a\hat{y}\} (y - \hat{y})$, for $a \neq 0$. Both of these families nest MSE ($k = 2$ in the former case and $a \to 0$ in the latter).*
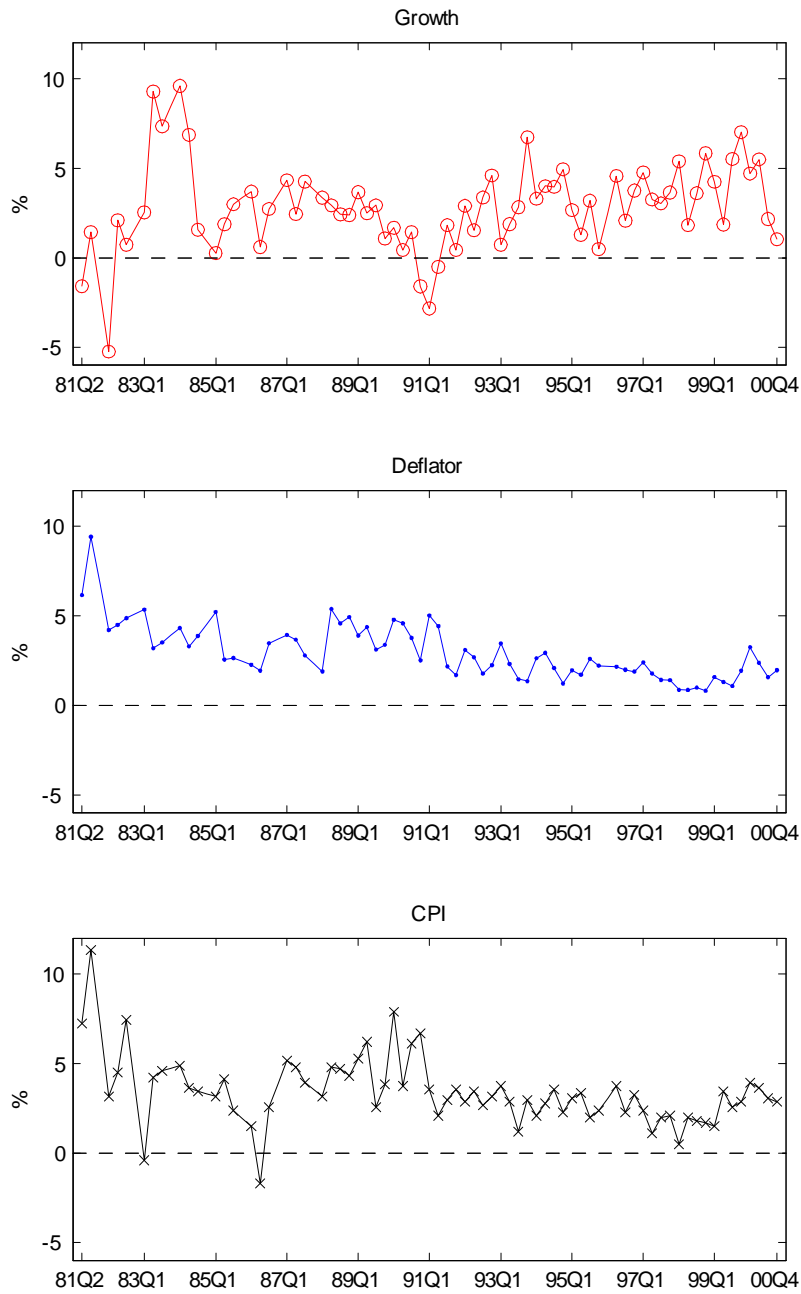
Figure 2: *Time series of annualized quarterly GDP growth, GDP deflator, and CPI inflation, over the period 1981Q2 to 2000Q4.*
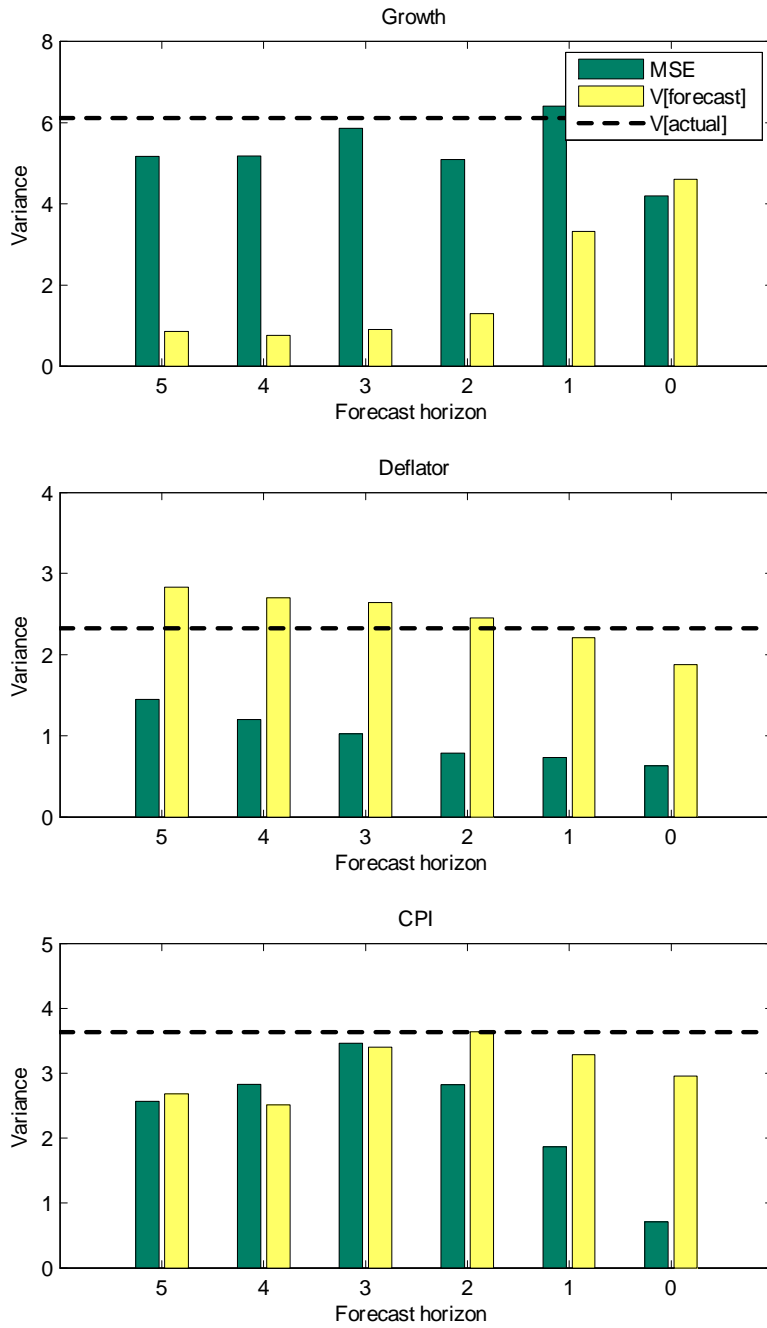
48

Figure 3: *Mean squared errors and forecast variances, for US GDP deflator, CPI inflation and GDP growth.*
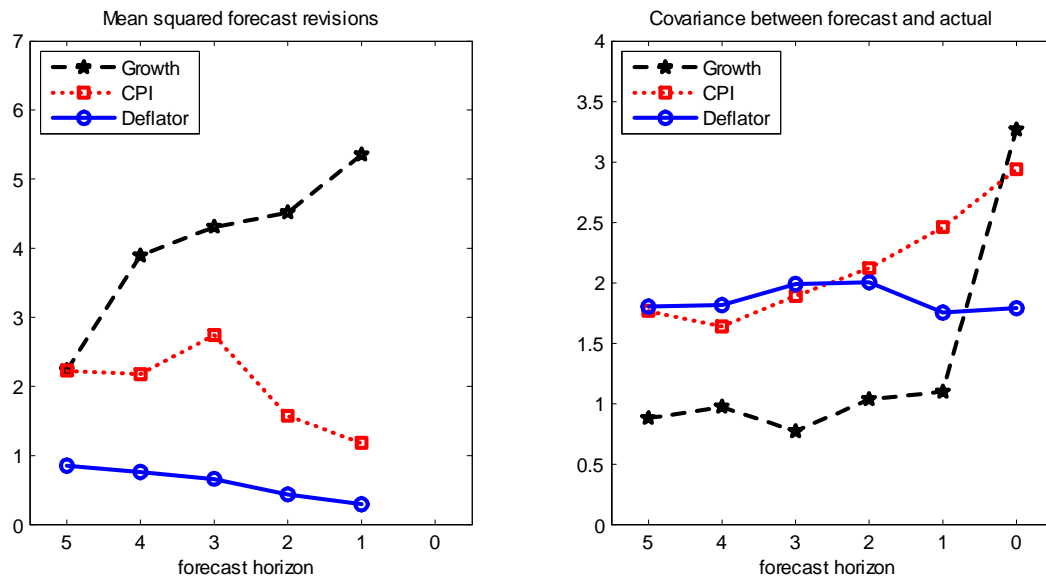
Figure 4: *Mean squared forecast revisions (left panel) and the covariance between forecasts and actuals, for US GDP deflator, CPI inflation and GDP growth.*