Forecast Rationality Tests Based on Multi-Horizon Bounds

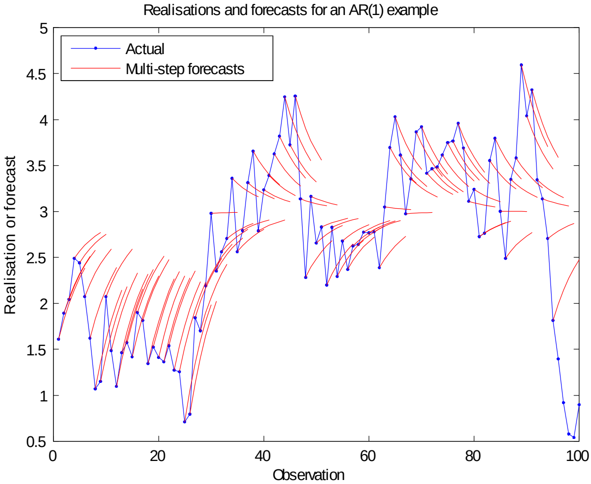
Andrew Patton and Allan Timmermann

Duke and UC-San Diego

4 November, 2010

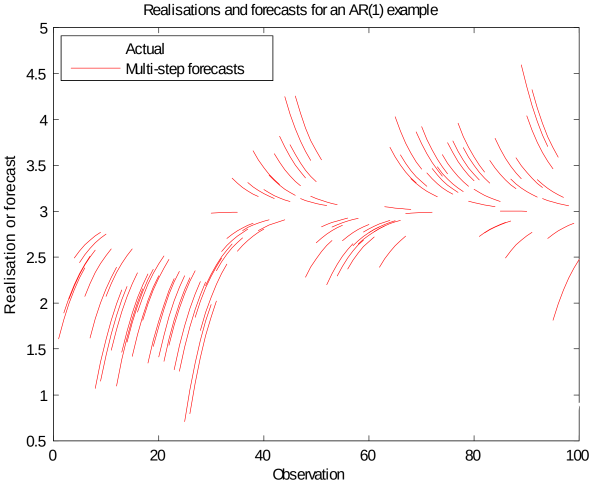
Actuals and forecasts for an AR(1) example

The "hedgehog" plot



Actuals and forecasts for an AR(1) example

The "hedgehog" plot, without data on the target variable



Background and motivation

- Under squared error loss, the optimal forecast is the conditional mean of the target variable (eg, Granger (1969)):

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

- The associated forecast error, $e^*_{t|t-h} \equiv Y_t - \hat{Y}^*_{t|t-h}$, should be mean-zero and orthogonal to any $Z_{t-h} \in \mathcal{F}_{t-h}$.

- Another implication of forecast optimality is that the MSE should be increasing in the horizon:

$$E\left[e^2_{t|t-1}\right] \leq E\left[e^2_{t|t-2}\right] \leq \cdots \leq E\left[e^2_{t|t-H}\right]$$

- Testing this property is complicated by the fact that it involves (potentially many) inequality constraints.

New tests of forecast rationality

- The availability of *multi-horizon* forecasts has created a need for tests of rationality that exploit information in the complete term structure of forecasts.
  - Surveys: SPF, ECB's SPF, Consensus Economics
  - Forecasts: OECD, IMF, Bank of England, Fed
  - Econometric models: Clements (1997), Marcellino, Stock and Watson (2006), Faust and Wright (2009)
- Multi-horizon forecasts allow us to draw more powerful conclusions about *joint* rationality, across all horizons.
- Joint rationality across all horizons requires the *internal consistency* of the set of forecasts, which can be tested without data on the target variable.

Contributions

- We propose several new tests of forecast rationality, exploiting information from multi-horizon forecasts
  - new monotonicity properties and second moment bounds that must hold across forecast horizons
  - generalized efficiency regressions that test the full sequence of forecast revisions
  - proposed tests are robust to certain forms of non-stationarity
- We present tests of rationality that may be used when the target variable is either not observable, or observable only with substantial measurement error.
- We study the size and power of these tests via simulations.
- We apply the tests to the Federal Reserve's "Greenbook" forecasts of inflation and GDP growth.

Outline of the talk

1. Background and motivation
2. **Testing forecast rationality across multiple horizons**
    1. Tests based on bounds
    2. Regression based tests
    3. Extensions
3. Simulation study of size and power
4. Evaluating the rationality of Greenbook forecasts
5. Summary and conclusions

Optimal forecasts under MSE (eg Granger, 1969)

Theorem 1: The optimal forecast under MSE loss

$$L\left(Y,\hat{Y}\right)=\left(Y-\hat{Y}\right)^2$$

is the conditional mean:

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

- Our paper explores additional implications of this well-known result and proposes empirical tests of these implications.
- Some of the results derived below make use of a standard covariance stationarity assumtion:

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

1a) Weakly increasing mean squared errors

- Forecast optimality under squared-error loss implies:

$$E_{t-h}\left[\left(Y_t - \hat{Y}^*_{t|t-h}\right)^2\right] \leq E_{t-h}\left[\left(Y_t - \tilde{Y}_{t|t-h}\right)^2\right] \text{ for any } \tilde{Y}_{t|t-h} \in \mathcal{F}_{t-h}$$

- In particular, a short-horizon forecast must be at least as good as a long-horizon forecast of the same variable:

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

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

by the law of iterated expectations.

1b) Weakly increasing mean squared forecast revisions

- Define a forecast revision as ($h_S < h_L$)

$$\begin{aligned} d^*_{t|h_S,h_L} &\equiv \hat{Y}^*_{t|t-h_S} - \hat{Y}^*_{t|t-h_L} \\ E\left[d^*_{t|h_S,h_L}\right] &= Cov\left[d^*_{t|h_S,h_L}, Z_{t-h_L}\right] = 0 \end{aligned}$$

- Mean squared forecast revisions are weakly increasing in h :

$$\begin{aligned} \text{Note that } d_{t|h_1,h_H} &\equiv \hat{Y}_{t|t-h_1} - \hat{Y}_{t|t-h_H} = \sum_{j=1}^{H-1} d_{t|h_j,h_{j+1}} \\ \text{so } \; V\left[d^*_{t|h_1,h_H}\right] &= V\left[\sum_{j=1}^{H-1} d^*_{t|h_j,h_{j+1}}\right] = \sum_{j=1}^{H-1} V\left[d^*_{t|h_j,h_{j+1}}\right] \\ \text{And thus } \; V\left[d^*_{t|h_1,h_2}\right] &\leq V\left[d^*_{t|h_1,h_3}\right] \leq \cdots \leq V\left[d^*_{t|h_1,h_H}\right] \end{aligned}$$

Weakly increasing MSE and MSR values (summary)

Corollary (1)

Under the assumptions of Theorem 1 and S1, it follows that, for any $h_S < h_M < h_L$,

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

and

$$V\left[d_{t|h_S,h_M}\right] \leq V\left[d_{t|h_S,h_L}\right].$$

Testing weakly increasing MSE

- Denote the MSE differentials (across horizons) as:

$$\Delta_j^e \equiv \mu_{h_j} - \mu_{h_{j-1}} = E\left[e_{t|t-h_j}^2\right] - E\left[e_{t|t-h_{j-1}}^2\right], \text{ for } j = 2, \ldots, H$$

Then rationality implies that:

$$\Delta^e \equiv [\Delta_2^e, \ldots, \Delta_H^e]' \geq \mathbf{0}.$$

- This can be tested via methods for handling multivariate inequality tests:

Multivariate inequality tests

- Early work: Bartholomew (1961), Kudo (1963), Perlman (1969), Gourieroux, et al. (1982) and Wolak (1987, 1989).
- Wolak test entertains (weak) monotonicity under the null hypothesis:

$$\begin{aligned} H_0 &: \Delta^e \geq \mathbf{0} \\ \text{vs. } H_1 &: \Delta^e \ngeq \mathbf{0} \end{aligned}$$

- These test statistics have a distribution under the null that 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 weights and $\chi^2(i)$ is a chi-squared variable with i degrees of freedom.
- Critical values are generally not known in closed form, but a set of approximate values can be calculated through Monte Carlo simulation (data-dependent weighted sum of chi-squared variables).

Multivariate inequality tests: bootstrap approaches

- Can convert the null and alternative

$$\begin{aligned} H_0 &: \Delta^e \geq \mathbf{0} \\ \text{vs.} \quad H_1 &: \Delta^e \ngeq \mathbf{0} \end{aligned}$$

into:

$$\begin{aligned} H_0 &: \min_j \Delta_j^e \geq 0 \\ \text{vs.} \quad H_1 &: \min_j \Delta_j^e < 0 \end{aligned}$$

- This can be tested using the bootstrap "reality check" of White (2000) and its extension by Hansen (2005).
    - May be particularly useful when the number of horizons (inequalities) is large.

2) Weakly decreasing mean squared forecasts

- Forecast rationality implies that the variance of the forecasts should be *decreasing* in the forecast horizon. Recall

$$Cov\left[e^*_{t|t-h}, Z_{t-h}\right] = 0 \Rightarrow Cov\left[\hat{Y}^*_{t|t-h}, e^*_{t|t-h}\right] = 0$$

$$\begin{aligned} \text{thus} \quad V\left[Y_t\right] &= V\left[\hat{Y}^*_{t|t-h}\right] + V\left[e^*_{t|t-h}\right] \\ \text{and so} \quad V\left[\hat{Y}^*_{t|t-h}\right] &= V\left[Y_t\right] - V\left[e^*_{t|t-h}\right] \end{aligned}$$

- Weakly increasing MSE implies a weakly decreasing variance of the forecast:

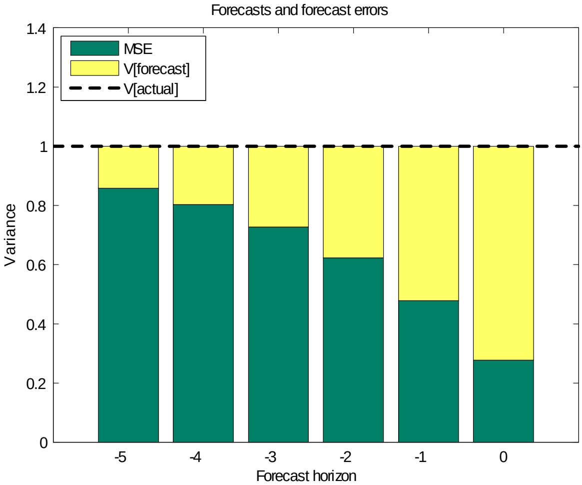
Corollary (2)

Under the assumptions of Theorem 1 and S1, we have

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

Weakly increasing MSE and weakly decreasing MSF

AR(1) example



Testing weakly decreasing MSF

- Denote the mean-squared forecast differentials (across horizons) as:

$$\Delta_h^f \equiv E\left[\hat{Y}_{t|t-h_j}^{*2}\right] - E\left[\hat{Y}_{t|t-h_{j-1}}^{*2}\right], \text{ for } j = 2, ..., H$$

  Then rationality implies that:

$$\Delta^f \equiv \left[\Delta_2^f, ..., \Delta_H^f\right]' \leq \mathbf{0}$$

- We can again test rationality through Wolak's (1989) test of the null and alternative hypotheses:

$$\begin{aligned} H_0 &: \Delta^f \leq \mathbf{0} \\ \text{vs. } H_1 &: \Delta^f \nleq \mathbf{0} \end{aligned}$$

MSF test: no need for data on the target variable

- This implication of multi-horizon forecast rationality can be tested *without data on the target variable*.
- Useful when the target variable is not observable (e.g., volatility) or subject to substantial measurement errors (some measures of economic growth - see, e.g., "real time" macro literature, Croushore, 2006
  - In this case the rationality test is a test of the *internal consistency* of the set of multi-horizon forecasts
  - This test has zero power to detect an internally-consistent set of forecasts that are totally independent of the stated target variable.

3) Weakly decreasing covariance btw forecast and target

- A further implication of decreasing forecast variances is that the covariance of the forecasts with the actuals should be *decreasing* in the forecast horizon:

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

Corollary (3)

Under the assumptions of Theorem 1 and S1, we obtain

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

Moreover,

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

4) Bounded variance of forecast revisions

- Forecast optimality implies:

$$V\left[e_{t|t-h_L}\right] \geq V\left[e_{t|t-h_S}\right] \quad \text{for } h_L > h_S$$

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

- Re-write short-horizon forecast as $\hat{Y}^*_{t|t-h_S} = \hat{Y}^*_{t|t-h_L} + d^*_{t|h_S,h_L}$, so:

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

$$\text{that is} \quad V\left[d^*_{t|h_S,h_L}\right] \leq 2Cov\left[Y_t, d^*_{t|h_S,h_L}\right]$$

Bounded variance of forecast revisions (cont)

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* $h_S < h_L$*. Then under the assumptions of Theorem 1 and S1, we have*

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

Moreover,

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

Bounded forecast revision variance: tests

- Bound on forecast revision:

$$\begin{aligned} V\left[d^*_{t|h_S,h_L}\right] &\leq 2Cov\left[Y_t, d^*_{t|h_S,h_L}\right] \quad \text{and so} \\ 0 &\leq E\left[2Y_t d^*_{t|h_S,h_L} - d^{*2}_{t|h_S,h_L}\right] \end{aligned}$$

  This limits the amount of variability in the forecast revisions, as a function of their covariance with the target variable.

- As above, this can be tested through

$$\begin{aligned} \Delta^b_h &\equiv E\left[2Y_t d_{t|h-1,h} - d^2_{t|h-1,h}\right], \text{ for } h = 2, ..., H \\ H_0 &: \Delta^b \geq \mathbf{0} \\ \text{vs. } H_1 &: \Delta^b \ngeq \mathbf{0} \end{aligned}$$

Summary of tests based on multi-horizon bounds

- The table below summarizes the relationships across horizons established in this paper, as $h \uparrow$:

	Y_t	$e^*_{t\|t-h}$	$\hat{Y}^*_{t\|t-h}$	$d^*_{t\|h_S,h_L}$
Y_t	σ^2_y	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
- Our analysis covers the remaining elements, with particular attention to cells (3,3) (3,4) and (4,4), which do not require data on the target variable.

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 above tests be able to detect it?

- We study this problem in two cases:

  1. Multi-step forecasts are obtained from a suite of horizon-specific models ("**direct**" multi-step forecasts)
  2. Forecasts for all horizons are obtained from a single model (and multi-step forecasts are obtained by "**iterating**" on the one-step model).

- We show via a simple example that the consistent use of a misspecified model may be detected using one of the mulit-horizon bounds presented in this paper.

Multi-horizon bounds - Direct forecasting

- Consider a target variable that evolves according to a stationary AR(2):

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

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

$$Y_t = \rho_h Y_{t-h} + v_t, \text{ for } h = 1, 2, \ldots$$

- By the properties of an AR(2) we have

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

- Eg, for $(\phi_1, \phi_2) = (0.1, 0.8)$ we find $\rho_1 = 0.5$ and $\rho_2 = 0.85$.

Multi-horizon bounds - Direct forecasting (cont'd)

- If $(\phi_1, \phi_2) = (0.1, 0.8)$ we find $\rho_1 = 0.5$ and $\rho_2 = 0.85$.
  1. This directly leads to a violation of the bound that the variance of the optimal forecast is *decreasing* in the horizon

  $$V\left[\hat{Y}_{t|t-1}\right] = \rho_1^2 \sigma_y^2 = 0.25\sigma_y^2 < V\left[\hat{Y}_{t|t-2}\right] = \rho_2^2 \sigma_y^2 = 0.72\sigma_y^2$$

  2. Further, it also violates the bound that the MSE from the optimal forecast is increasing in the horizon:

  $$\begin{aligned} MSE_1 &\equiv E\left[\left(Y_t - \hat{Y}_{t|t-1}\right)^2\right] = \sigma_y^2\left(1 - \rho_1^2\right) = 0.75\sigma_y^2, \\ MSE_2 &\equiv E\left[\left(Y_t - \hat{Y}_{t|t-2}\right)^2\right] = \sigma_y^2\left(1 - \rho_2^2\right) = 0.28\sigma_y^2. \end{aligned}$$

- The forecaster should recognize that the 2-step forecast is better than the 1-step forecast, and so simply use the 2-step forecast again for the 1-step forecast.
  - Or better yet, improve the forecasting models being used..

Multi-horizon bounds - Iterated forecasting

- Consider again a target variable that evolves according to a stationary AR(2) as above, but the forecaster uses an AR(1) model:

$$\begin{aligned} Y_t &= \rho_1 Y_{t-1} + v_t, \\ \text{so} \quad \hat{Y}_{t|t-h} &= \rho_1^h Y_{t-h}, \text{ for } h = 1, 2, \ldots \end{aligned}$$

  where $\rho_1 = \phi_1 / (1 - \phi_2)$ is the population value of the AR(1) parameter when the DGP is an AR(2).

- The MSEs for the $h = 1$ and $h = 2$ forecasts are:

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

- If $(\phi_1, \phi_2) = (0.1, 0.8)$, then

$$MSE_1 = 0.75\sigma_y^2 > MSE_2 = 0.6375\sigma_y^2.$$

- Thus the MSE bound is violated, and a test based on this bound would detect the use of a misspecified model.

Multi-horizon bounds - Iterated forecasting (cont'd)

- In fact, a bound based *solely on the forecasts* may also detect this model misspecification. Consider the MSFR for this example:

$$\begin{aligned} MSFR_{1,2} &\equiv E\left[\left(\hat{Y}_{t|t-1} - \hat{Y}_{t|t-2}\right)^2\right] = \rho_1^2\sigma_y^2\left(1-\rho_1^2\right) \\ MSFR_{1,3} &\equiv E\left[\left(\hat{Y}_{t|t-1} - \hat{Y}_{t|t-3}\right)^2\right] = \rho_1^2\sigma_y^2\left(1+\rho_1^2-2\rho_1^2\rho_2\right) \end{aligned}$$

- If $(\phi_1, \phi_2) = (0.1, 0.8)$, then $\rho_1 = 0.5$ and $\rho_2 = 0.85$ and

$$MSFR_{1,2} = 0.19\sigma_y^2 > MSFR_{1,3} = 0.16\sigma_y^2.$$

- This is further evidence of model misspecification in this example.

Outline of the talk

1. Background and motivation
2. Testing forecast rationality across multiple horizons
    1. Tests based on bounds
    2. **Regression based tests**
    3. Extensions
3. Simulation study of size and power
4. Evaluating the rationality of Greenbook forecasts
5. Summary and conclusions

Regression-based tests of forecast rationality

- We also consider Mincer-Zarnowitz forecast rationality regressions:

$$Y_t = \alpha_h + \beta_h \hat{Y}_{t|t-h} + v_{t|t-h}. \tag{1}$$

Corollary (5)

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

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

- One approach, adopted in Capistrán (2007) is to run MZ regressions for each horizon, and then use Bonferroni bounds to obtain a joint test:
  1. Obtain the p-values from a chi-squared test (with 2 degrees of freedom) of the above null for each horizon
  2. Reject rationality 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.

Vector MZ tests

- Stack the individual MZ regressions into a system and estimate them jointly:

$$\begin{bmatrix} Y_{t+1} \\ \vdots \\ Y_{t+H} \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_H \end{bmatrix} + \begin{bmatrix} \beta_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \beta_H \end{bmatrix} \begin{bmatrix} \hat{Y}_{t+1|t} \\ \vdots \\ \hat{Y}_{t+H|t} \end{bmatrix} + \begin{bmatrix} u^1_{t+1} \\ \vdots \\ u^H_{t+H} \end{bmatrix}$$

$$\begin{aligned} H_0 &: \quad \alpha_1 = ... = \alpha_H = 0 \cap \beta_1 = ... = \beta_H = 1 \\ H_1 &: \quad \alpha_1 \neq 0 \cup ... \cup \alpha_H \neq 0 \cup \beta_1 \neq 1 \cup ... \cup \beta_H \neq 1 \end{aligned}$$

- The vector residual of this regression will, even under the null of rationality, exhibit some autocorrelation (and possibly heteroskedasticity) and so robust standard errors must be obtained.

A univariate optimal revision test

- Combining the MZ regression approach with the representation of a short-horizon forecast as a function of a long-horizon forecast and the intervening forecast revisions suggests a test of forecast rationality:

$$Y_t = \alpha + \beta_H \hat{Y}_{t|t-h_H} + \sum_{j=1}^{H-1} \beta_j d_{t|h_j,h_{j+1}} + u_t. \qquad (2)$$

Corollary

Under the assumptions of Theorem 1 and S1, the population values of the parameters in the optimal revision regression in equation (2) satisfy

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

Regression-based tests w/o the target variable

$$\hat{Y}_{t|t-1} = \tilde{\alpha}_h + \tilde{\beta}_h \hat{Y}_{t|t-h} + \tilde{u}_{t|t-h} \quad (3)$$

$$\hat{Y}_{t|t-h_1} = \tilde{\alpha} + \tilde{\beta}_H \hat{Y}_{t|t-h_H} + \sum_{j=2}^{H-1} \tilde{\beta}_j d_{t|h_j,h_{j+1}} + v_t, \quad (4)$$

Corollary

Under the assumptions of Theorem 1 and S1, the population values of the parameters in (a) Mincer-Zarnowitz regression by proxy in equation (3) satisfy

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

and (b) the population values of the parameters in the optimal revision regression by proxy, in equation (4) satisfy

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

Extensions of the main results

- We now present two extensions of the main results:

1. Heterogeneity in the **data**
2. Heterogeneity in the **forecast horizons**

- The first of these extensions is motivated by empirical evidence of structural breaks in the variance of some macroeconomic series (eg, the Great Moderation)
- The second is more practical: sometimes we need to aggregate forecasts into "buckets" of short and long horizons rather than treat all unique horizons separately.

Allowing for heterogeneity in the data

- By optimality, we have

$$
\begin{aligned}
E_t\left[\left(Y_{t+h}-\hat{Y}^*_{t+h|t}\right)^2\right] &\leq E_t\left[\left(Y_{t+h}-\hat{Y}^*_{t+h|t-j}\right)^2\right] \quad \text{for } j>0 \\
\text{so}\quad E\left[\left(Y_{t+h}-\hat{Y}^*_{t+h|t}\right)^2\right] &\leq E\left[\left(Y_{t+h}-\hat{Y}^*_{t+h|t-j}\right)^2\right] \quad \text{by the LIE}
\end{aligned}
$$

- Under stationarity this implies

$$
E\left[\left(Y_{t+h}-\hat{Y}^*_{t+h|t}\right)^2\right] \leq E\left[\left(Y_{t+h+j}-\hat{Y}^*_{t+h+j|t}\right)^2\right]
$$

- But if we use a *fixed event framework* we do not need stationarity, and we focus on the second line above
  - That is, keep the target variable (Y_{t+h}) fixed, and vary the horizon of the forecast (from $t-j$ to t).

Allowing for heterogeneity in the data, cont'd

Define the following variables

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

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

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

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

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

Allowing for heterogeneity in the data

Proposition

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

$$
\begin{array}{ll}
(a) & \overline{MSE}_T\,(h_S) \le \overline{MSE}_T\,(h_L) \\
(b) & \overline{MSF}_T\,(h_S) \ge \overline{MSF}_T\,(h_L) \\
(c) & \overline{C}_T\,(h_S) \ge \overline{C}_T\,(h_L) \\
(d) & \overline{MSFR}_T\,(h_S, h_M) \le \overline{MSFR}_T\,(h_S, h_L) \\
(e) & \overline{MSFR}_T\,(h_S, h_L) \le 2\overline{B}_T\,(h_S, h_L)
\end{array}
$$

Allowing for heterogeneity in the data: discussion

- Allowing for heterogeneity in the data need not affect the bounds obtained 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 above 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):

Tests under data heterogeneity (brief)

Proposition

Define

$$
\begin{aligned}
d_{jt}^{MSE} &\equiv \left(Y_t - \hat{Y}_{t|t-h_j}^{*}\right)^2 - \left(Y_t - \hat{Y}_{t|t-h_{j-1}}^{*}\right)^2, \ \ \textit{for } j = 2, ..., H \\
d_{jt}^{MSF} &\equiv \hat{Y}_{t|t-h_j}^{*2} - \hat{Y}_{t|t-h_{j-1}}^{*2}, \ \ \textit{for } j = 2, ..., H \\
&\vdots \\
\mathbf{d}_t^k &\equiv \left[d_{qt}^k, ..., d_{Ht}^k\right]', \ \ \hat{\Delta}_T^k \equiv \frac{1}{T}\sum_{t=1}^{T} \mathbf{d}_t^{k\prime}, \ \ V_T \equiv V\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T} \mathbf{d}_t^{k\prime}\right],
\end{aligned}
$$

where $k \in \{MSE, MSF, C, MSFR, B)$. *Assume: (i)* $\mathbf{d}_t^k = \Delta^k + \epsilon_t^k$, *for* $t = 1, 2, ...,\ \Delta \in \mathbb{R}^{H-1}$; *(ii)* ϵ_t^k *is strong mixing of size* $-r/(r-2)$, $r > 2$; *(iii)* $E\left[\epsilon_t^k\right] = 0$ *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. Then:*

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

Allowing for heterogeneous forecast 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.
  - Eg, 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 hetergeneous horizons.
  - The key to this proposition is that any "short" horizon forecast must have a corresponding "long" horizon forecast.

Allowing for heterogeneous forecast horizons

Proposition

Consider a data set of the form $\{(Y_t, \hat{Y}^*_{t|t-h_t}, \hat{Y}^*_{t|t-h_t-k_t})\}_{t=1}^{T}$, *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:*

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

Outline of the talk

1. Background and motivation
2. Testing forecast rationality across multiple horizons
   1. Tests based on bounds
   2. Regression based tests
   3. Extensions
3. **Simulation study of size and power**
4. Evaluating the rationality of Greenbook forecasts
5. Summary and conclusions

Simulation design: DGP and Optimal forecasts

- Simple AR(1) is the DGP:

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

- Calibrate the parameters to quarterly US CPI inflation data:

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

- Optimal forecasts are the conditional expectation:

$$\begin{aligned} \hat{Y}^*_{t|t-h} &= E_{t-h}\left[Y_t\right] \\ &= \mu_y + \phi^h \left(Y_{t-h} - \mu_y \right) , \;\; \text{for } h = 1, 2, ..., H \\ H &\in \{ \, 4 \; , \; 8 \, \} \end{aligned}$$

Simulation design: Measurement error

- The measured value of the target variable is assumed to be subject to error:

$$\begin{aligned} \tilde{Y}_t &= Y_t + \eta_t \\ \text{with } \eta_t &\sim iid\ N\left(0, \sigma_\eta^2\right) \end{aligned}$$

- Magnitude of the measurement error:

  1. *zero* (as for CPI)
  2. *medium* (as for GDP growth first release)
  3. *high* (twice the *medium* value):

$$\sigma_\eta^2 / \sigma_y^2 \in \{\ 0\ ,\ 0.70\ ,\ 1.40\ \}$$

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

Simulation design: Sub-optimal forecasts

1. Forecasts subject to equal noise at all horizons:

$$\hat{Y}_{t|t-h} = \hat{Y}^{*}_{t|t-h} + \sigma_{\xi,h} Z_{t,t-h}, \text{ where } Z_{t,t-h} \sim iid\ N(0,1)$$
$$\sigma_{\xi,h} = \sqrt{0.70}\sigma_y \ \ \forall\ h$$

2. Forecasts subject to noise *increasing* in the horizon:

$$\sigma_{\xi,h} = \frac{2(h-1)}{H-1} \times \sqrt{0.70}\sigma_y, \text{ for } h = 1, 2, ..., H$$

Simulation results: Size of inequality-based tests

Finite-sample size is lower than nominal size of 10%

	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

Simulation results: Size of regression-based tests

MZ-Bonf is over-sized, Vector MZ does terribly, MZ-revisions do nicely.

	H = 4			H = 8		
Meas. error variance:	*High*	*Med*	*Zero*	*High*	*Med*	*Zero*
MZ on short horizon	10.8	11.9	13.6	10.8	11.9	13.6
Opt. rev reg.	10.2	9.7	9.8	11.5	10.2	9.4
Opt. rev reg., with proxy	10.8	10.8	10.8	9.5	9.5	9.5
Univar MZ, Bonf	12.5	12.9	18.2	18.4	19.1	22.4
Univar MZ, Bonf, 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

Simulation results: Power against equal noise across h

	H = 4			H = 8		
Meas. error variance:	*High*	*Med*	*Zero*	*High*	*Med*	*Zero*
Inc MSE	7.1	6.5	5.0	15.8	14.4	13.2
Dec COV	6.0	5.1	4.9	14.9	13.8	13.0
COV bound	72.4	78.0	82.5	73.5	78.9	84.0
Dec MSF	6.0	6.0	6.0	18.2	18.2	18.2
Inc MSFR	8.1	8.1	8.1	16.7	16.7	16.7
Dec COV, with proxy	8.4	8.4	8.4	15.5	15.5	15.5
COV bound, with proxy	98.5	98.5	98.5	99.2	99.2	99.2
MZ on short horizon	92.6	98.0	100.0	92.6	98.0	100.0
Opt. revision regr.	84.4	94.0	99.6	73.9	88.0	99.0
Opt. revision regr., proxy	100.0	100.0	100.0	100.0	100.0	100.0
Bonf, using actuals	68.4	83.8	97.7	67.3	79.3	95.4
Bonf, using forecasts only	100.0	100.0	100.0	100.0	100.0	100.0
Bonf, all tests	100.0	100.0	100.0	100.0	100.0	100.0

Simulation results: Power against noise increasing in h

	H = 4			**H = 8**		
Meas. error variance:	*High*	*Med*	*Zero*	*High*	*Med*	*Zero*
Inc MSE	0.5	0.2	0.1	0.1	0.4	0.1
Dec COV	3.4	3.4	4.0	12.1	11.2	12.5
COV bound	13.3	14.6	16.0	89.1	91.6	96.0
Dec MSF	45.2	45.2	45.2	100.0	100.0	100.0
Inc MSFR	0.0	0.0	0.0	0.0	0.0	0.0
Dec COV, with proxy	5.0	5.0	5.0	13.9	13.9	13.9
COV bound, with proxy	72.7	72.7	72.7	100.0	100.0	100.0
MZ on short horizon	10.8	11.9	13.6	10.8	11.9	13.6
Opt. revision regr.	9.0	8.6	11.0	9.3	9.9	11.6
Opt. revision regr., proxy	66.1	66.1	66.1	53.9	53.9	53.9
Bonf, using actuals	12.7	12.2	12.8	99.4	99.4	99.5
Bonf, using forecasts only	63.9	63.9	63.9	100.0	100.0	100.0
Bonf, all tests	52.3	52.1	52.0	100.0	100.0	100.0

Outline of the talk

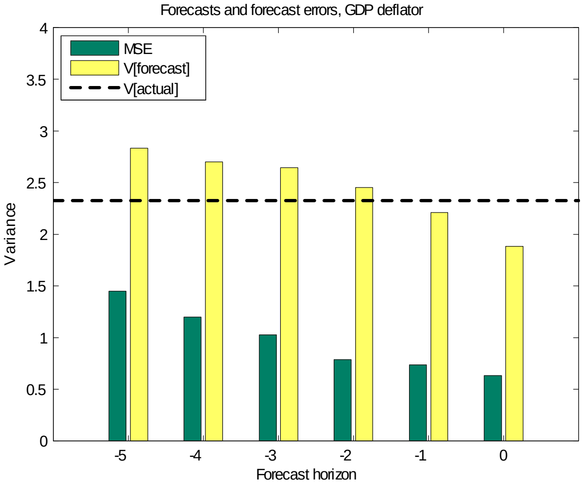
1. Background and motivation
2. Testing forecast rationality across multiple horizons
    1. Tests based on bounds
    2. Regression based tests
    3. Extensions
3. Simulation study of size and power
4. **Evaluating the rationality of Greenbook forecasts**
5. Summary and conclusions

Evaluating the rationality of the "Greenbook" forecasts

- We study the Fed's "Greenbook" forecasts of the GDP growth, GDP deflator and CPI inflation.
  - Data are quarterly, over the period 1982Q1 to 2000Q4, approx. 80 observations
  - Data are from Faust and Wright (2009), who constructed the Greenbook forecasts and actuals from real-time Fed publications. We have aligned these in "event time" to fit the structure assumed by our theory.
- We have 6 forecast horizons: $h = 0, 1, 2, 3, 4, 5$

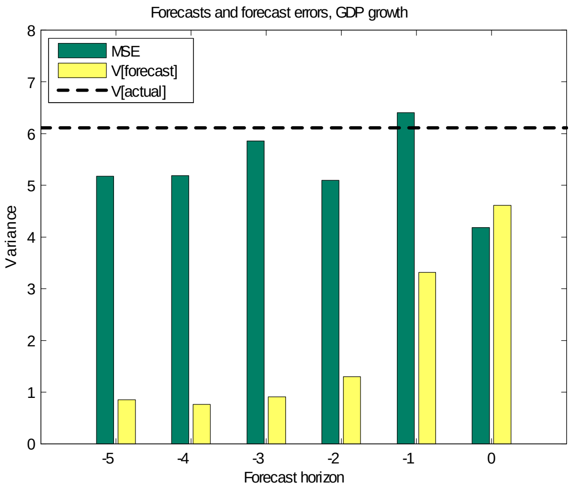
Increasing MSE and decreasing MSF

Greenbook forecasts of GDP deflator, 1982Q1-2000Q4



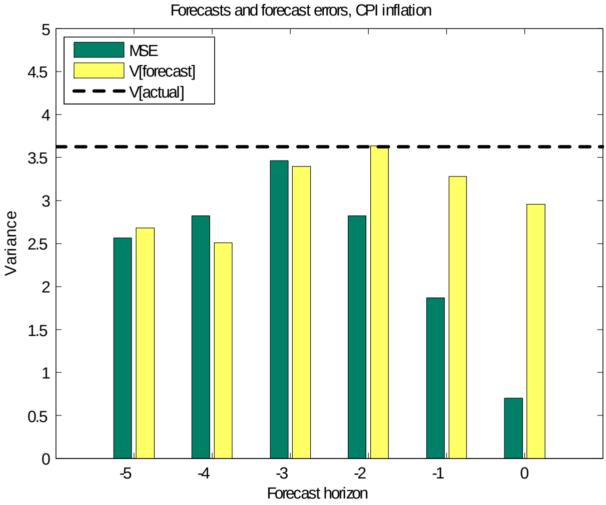
Increasing MSE and decreasing MSF

Greenbook forecasts of GDP growth, 1982Q1-2000Q4



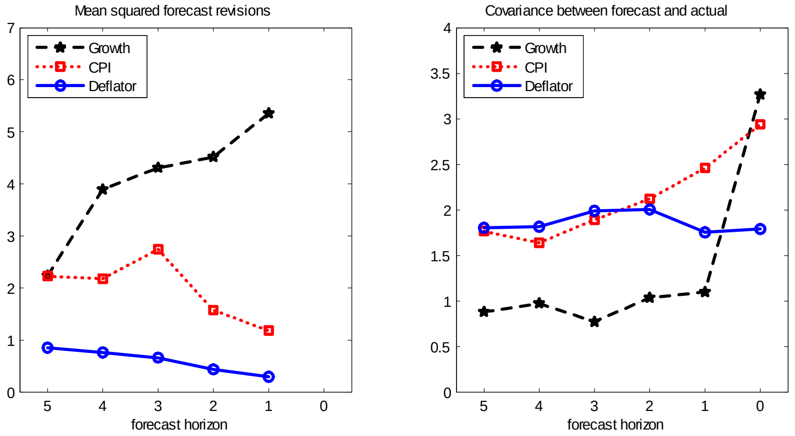
Increasing MSE and decreasing MSF

Greenbook forecasts of CPI inflation, 1982Q1-2000Q4



Increasing MSFR and decreasing Cov[yhat,y]

Greenbook forecasts, 1982Q1-2000Q4



P-values from tests of multi-horizon forecast rationality

#	*Series:*	*Growth*	*Deflator*	*CPI*
1	Inc MSE	0.591	0.966	0.639
2	Dec COV	0.879	**0.057***	0.991
3	COV bound	0.560	**0.000***	**0.009***
4	Dec MSF	0.916	**0.026***	0.719
5	Inc MSFR	**0.089***	0.938	0.620
6	Dec COV, with proxy	0.807	**0.075***	0.772
7	COV bound, with proxy	0.206	**0.010***	0.656
8	MZ on short horizon	0.245	0.313	0.699
9	Opt. rev regr.	0.709	**0.000***	**0.001***
10	Opt. rev regr., with proxy	**0.000***	**0.009***	**0.022***
11	Bonf, using actuals	1.000	**0.000***	**0.004***
12	Bonf, using forecasts only	**0.000***	**0.047***	0.108
13	**Bonf, all tests**	**0.000***	**0.001***	**0.010***

Interpretation of the test results

- **Growth:** We find a strong rejection of internal consistency via the "Optimal revision" test, and a mild violation of the increasing mean-squared forecast revision test.
- **GDP deflator:** We find many violations, of decreasing covariance, the covariance bound on forecast revisions, decreasing mean-squared forecast, and through the optimal revision regressions.
- **CPI inflation:** We find a violation of the bound on the variance of the revisions, and a rejection through the optimal revision regressions
- **Overall:** In all cases the Bonferroni-based combination test rejects multi-horizon forecast rationality at the 0.05 level. The sources of the rejections give some clues as to possible sources of sub-optimality.

Summary and conclusions

- We propose new tests of forecast rationality exploiting information from multi-horizon forecasts
  - Forecast rationality implies bounds on second moments across forecast horizons
  - Can be tested via tests of moment inequalities
  - Some of these bounds can be tested *without* data on the target variable
  - Bounds hold in the presence of heterogeneous data and heterogeneous horizons
- Simulation results indicate that the new tests are somewhat conservative, but retain power in scenarios where extant tests are weak
- Applying the new tests to the Fed's Greenbook forecasts, we find evidence against rationality for all three variables
  - Forecasts of the GDP deflator seem particularly bad
  - Forecasts of growth do not reject rationality when data on the "actual" is used, but tests using only data on the forecasts do reject