Journal of Business & Economic Statistics
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/ubes20

Comment

Kajal Lahiri [a]
[a] Department of Economics, University at Albany: SUNY, Albany, NY, 12222

Available online: 22 Feb 2012

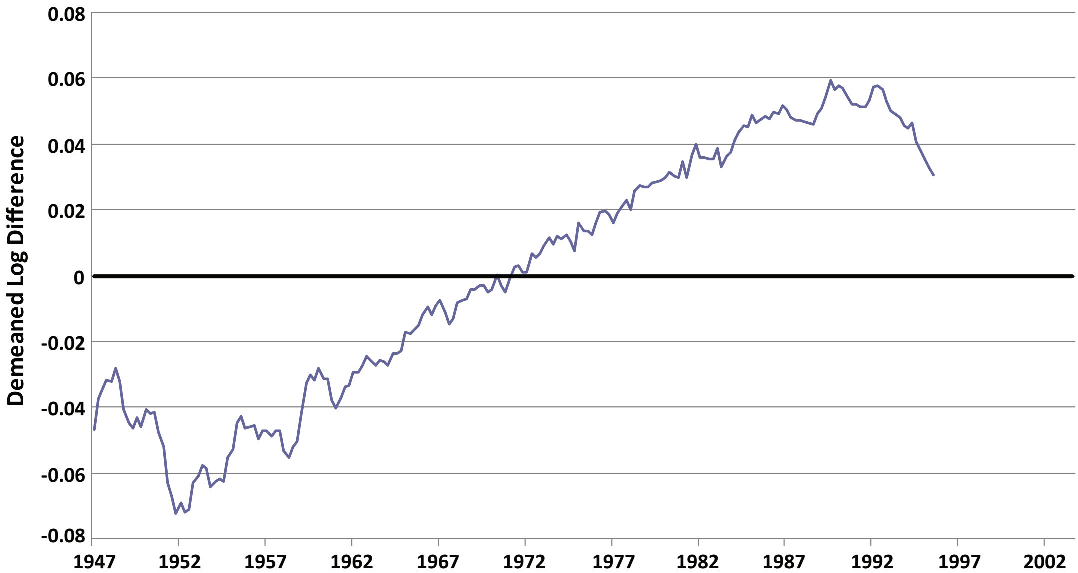

Figure 3. Stark plot across January 1996 benchmark revision. The plot shows the demeaned log differences of GDP before and after the benchmark revision of January 1996. It is a plot of $\log[X(t,b)/X(t,a)] - m$, where $X(t,s)$ is the level of X at date t from vintage s, where $s = a$ or $s = b$, $b > a$, and m is the mean of $\log[X(\tau,b)/X(\tau,a)]$ for all the dates that are common to both vintages a and b. In this plot, a = December 1995 and b = October 1999. (Color figure available online.)

than other tests, including the standard Mincer–Zarnowitz test and the test for zero-mean forecast errors.

To conclude, this article by Patton and Timmermann provides us with an excellent set of tests that can complement much existing research. The tests help us cross two dimensions of forecast rationality: horizon and real-time vintage. They could potentially help as well in the subsample dimension.

[Received April 2011. Revised August 2011.]

REFERENCES

Brown, B. W., and Maital, S. (1981), "What Do Economists Know? An Empirical Study of Experts' Expectations," *Econometrica*, 49, 491–504. [17]

Croushore, D. (2010), "An Evaluation of Inflation Forecasts From Surveys Using Real-Time Data," *B.E. Journal of Macroeconomics: Contributions*, 10, 10. [17]

——— (2011), "Two Dimensions of Forecast Analysis," Working Paper, University of Richmond. [17]

Croushore, D., and Stark, T. (2001), "A Real-Time Data Set for Macroeconomists," *Journal of Econometrics*, 105, 111–130. [18,19]

Faust, J., Rogers, J. H., and Wright, J. H. (2005), "News and Noise in G-7 GDP Announcements," *Journal of Money, Credit, and Banking*, 37, 403–419. [19]

Keane, M. P., and Runkle, D. E. (1990), "Testing the Rationality of Price Forecasts: New Evidence From Panel Data," *American Economic Review*, 80, 714–735. [17]

Patton, A. J., and Timmermann, A. (2011), "Forecast Rationality Tests Based on Multi-Horizon Bounds," *Journal of Business and Economic Statistics*, this issue. [18]

Zarnowitz, V. (1985), "Rational Expectations and Macroeconomic Forecasts," *Journal of Business & Economic Statistics*, 3, 293–311. [17]

Comment

Kajal Lahiri
Department of Economics, University at Albany: SUNY, Albany, NY 12222
(*klahiri@albany.edu*)

1. INTRODUCTION

I enjoyed reading yet another article by Patton and Timmermann (PT hereafter) and feel that it has broken new ground in testing the rationality of a sequence of multi-horizon fixed-target forecasts. Rationality tests are not new in the forecasting literature, but the idea of testing the monotonicity properties of second moment bounds across several horizons is novel and can suggest possible sources of forecasting failure. The basic premise is that since fixed-target forecasts at shorter horizons are based on more information, they should on the average be more accurate than their longer horizon counterparts. The internal consistency properties of squared errors, squared forecasts, squared forecast revisions, and the covariance between the target variable and the forecast revision are tested as inequality

Journal of Business & Economic Statistics
January 2012, Vol. 30, No. 1
DOI: 10.1080/07350015.2012.634342

constraints across horizons. They also generalize the single-horizon Mincer-Zarnowitz (MZ) unbiasedness test by estimating a univariate regression of the target variable on the longest horizon forecast and all intermediate forecast revisions. Using a Monte Carlo experiment and Greenbook forecasts of four macro variables, PT show that the covariance bound test and the generalized MZ regression using all interim forecast revisions have good power to detect deviations from forecast optimality. I am sure we will be using, extending, and finding caveats with some of the testing proposals suggested in this article for years to come.

2. THEORETICAL CONSIDERATIONS

An important starting point of the article is that for internal consistency of a sequence of optimal forecasts, the variance of forecasts should be a weakly decreasing function of the forecast horizon. This point has been discussed by Isiklar and Lahiri (2007), and was originally the basis for testing whether data revisions are news (and not noise) by Mankiw and Shapiro (1986). In order to highlight the nature of the fixed-target forecast variances, I have plotted in Figure 1 the sequences of monthly real GDP forecasts for 24 target years 1986–2009 from horizons 24 to 1 using consensus forecasts from the Blue Chip surveys. This survey is very well suited to examining the dynamics of forecasts over horizons. The respondents start forecasting in January of the previous year, and their last forecast is reported at the beginning of December of the target year. The shaded bars in the bottom of Figure 1 are the variances of mean forecasts calculated over the target years. Clearly, the variances are nondecreasing functions of horizons and thus the relationship is consistent with rational expectations. Isiklar and Lahiri (2007) explained the relationship by the following logic. Consider $y_t = f_{t,h} + u_{t,h}$ where y_t is the actual GDP growth, $f_{t,h}$ is the h-period ahead forecast ($h = 24, 23, \ldots, 1$) made at time $t - h$, and $u_{t,h}$ denotes the ex-post error associated with this forecast. Since rational expectations imply that $\mathrm{cov}(f_{t,h}, u_{t,h}) = 0$, we have $\mathrm{var}(y_t) = \mathrm{var}(f_{t,h}) + \mathrm{var}(u_{t,h})$, which implies (since for fixed target forecasts, variance of y_t is same for all h) that the variations in forecasts and forecast errors move in opposite directions as the forecast horizon changes. Therefore, as the forecast horizon decreases, the forecast error variability (and therefore the uncertainty) also decreases, but the forecast variability increases. Another way of looking at this increasing variability of forecasts is that as the forecast horizon decreases, more information is absorbed in the forecasts, thus increasing their variability. This information accumulation process can be seen using a simple moving average (MA) data-generating process. Suppose that the actual process has a moving average representation of order q so that $y_t = \mu + \sum_{k=0}^{q} \theta_k \varepsilon_{t-k}$ with $\mathrm{var}(\varepsilon_t) = \sigma^2$. Let $I_{t,h}$ denote the information available at time t-h. Then, the optimal forecast at horizon h will be

$$f_{t,h} \equiv E(y_t | I_{t,h}) = \mu + \sum_{k=h}^{q} \theta_k \varepsilon_{t-k}, \tag{1}$$

and the variance of the forecast is

$$\mathrm{var}(E(y_t | I_{t,h})) = \sigma^2 \sum_{k=h}^{q} \theta_k^2. \tag{2}$$

Similarly, the variance of the forecast when the forecast horizon is $h - 1$ is $\mathrm{var}(E(y_t | I_{t,h-1})) = \sigma^2 \sum_{k=h-1}^{q} \theta_k^2$, so that $\mathrm{var}(f_{t,h-1}) = \mathrm{var}(f_{t,h}) + \theta_{h-1}^2 \sigma^2$.

Thus, when the forecast horizon is very long, that is, several years, the forecasts tend to converge toward the mean of the process, and as information is accumulated, the forecasts change increasing the forecast variability. Figure 1 exhibits this phenomenon very well. Note that for horizons from 24 to 16, the variance seems to remain constant, as was illustrated by Isiklar and Lahiri (2007) for a large number of countries, but with a smaller sample size using the Consensus Survey forecasts. The forecast variability increases because of the variability of the accumulated shocks, that is, $\theta_k \varepsilon_{t-k}$. Therefore, if forecast variability does not change over several long horizons, this may mean that the information acquired at 24 to 16 horizons does not have much impact on the actual value, that is, $|\theta_k|$ is small or equivalently relevant information simply does not exist. Of course, this may also be related to the informational inefficiency of the forecasts. It is possible that even if potentially relevant information over these horizons were available, the forecasters did not incorporate them appropriately causing less than optimal variability in the forecasts. The point here is that due to the nonmonotone arrival and use of information by forecasters at different horizons, the monotonicity properties of second moments like the forecast variance that PT exploit may be less obliging for the detection of forecast suboptimality.

The first difference in the MSE_h provides a measure of the new information content of forecasts when the horizon is h. On the basis of Equation (1), an optimal forecast $f_{t,h}$ satisfies

$$\Delta\mathrm{MSE}(f_{t,h}) \equiv \mathrm{MSE}(f_{t,h+1}) - \mathrm{MSE}(f_{t,h}) = \theta_h^2 \sigma^2, \tag{3}$$

which is equivalent to the information content of the new information in the actual process.

Now let $\tilde{I}_{t,h}$ denote a strict subset of $I_{t,h}$, and $\tilde{f}_{t,h}$ be a suboptimal forecast, which is generated according to

$$\tilde{f}_{t,h} \equiv E(y_t | \tilde{I}_{t,h}) = \tilde{\mu} + \sum_{k=h}^{q} \tilde{\theta}_k \tilde{\varepsilon}_{t-k}, \tag{4}$$

where q denotes the longest forecast horizon at which the first fixed-target forecast is reported—it defines the conditional mean of the actual process when the horizon is q, that is, $\tilde{\mu} = E(y_t | \tilde{I}_{t,q})$; $\tilde{\varepsilon}_{t-h}$ denotes the "news" component used by the forecaster, and $\tilde{\theta}_h$ denotes the impact of this news component as perceived by the forecaster.

For convenience, let us assume that the forecasters observe the news ε_{t-h} correctly, but that their utilization of news is not optimal, so that $\tilde{\theta}_h \neq \theta_h$ and $\tilde{\varepsilon}_{t-h} = \varepsilon_{t-h}$. Thus, we see that the forecast errors follow:

$$y_t - \tilde{f}_{t,h} = (\mu - \tilde{\mu}) + \sum_{k=h}^{q} (\theta_k - \tilde{\theta}_k)\varepsilon_{t-k} + \sum_{k=0}^{h-1} \theta_k \varepsilon_{t-k}, \tag{5}$$

where the first component on the right-hand side (RHS) denotes the bias in the forecast, the second component denotes the error due to inefficiency, and the third component denotes the error due to unforecastable events after the forecast is reported. Calculating mean squared error (MSE) and assuming that sample estimates converge to their population values, we

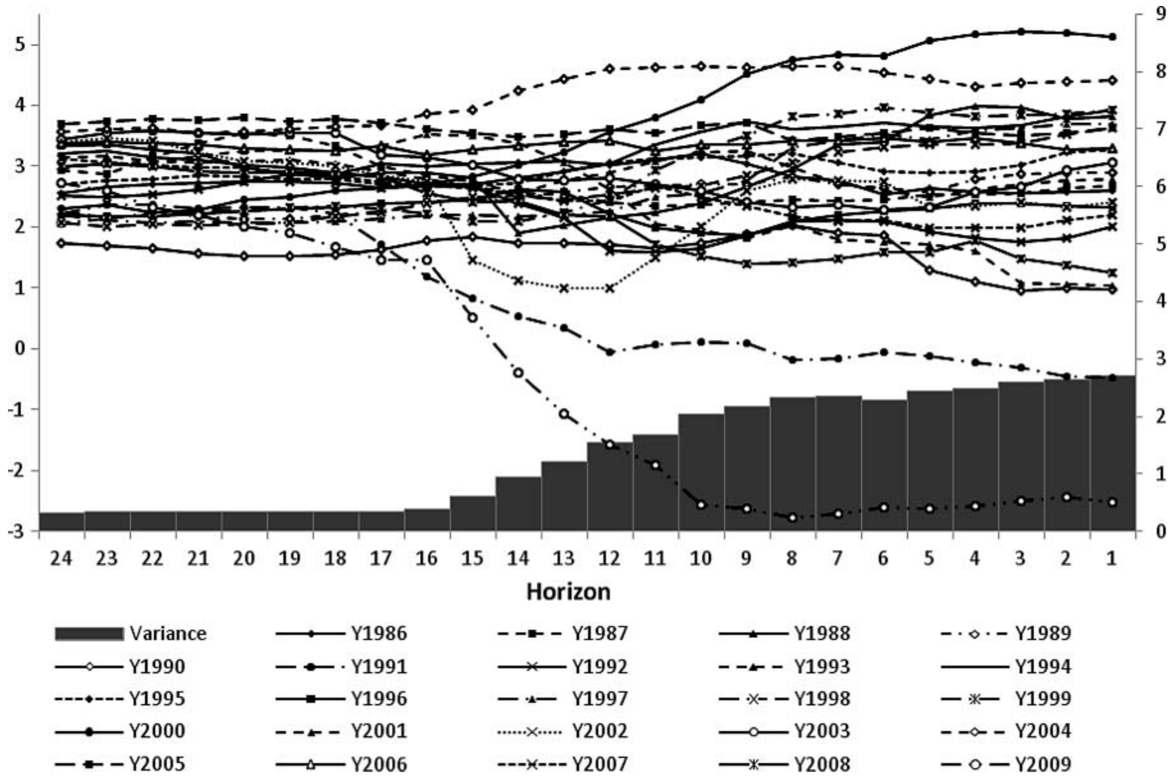

Figure 1. Evolution of fixed target forecasts over horizons and their variances.

get

$$\mathrm{MSE}_h = (\mu - \tilde{\mu})^2 + \sum_{k=h}^{q} (\theta_k - \tilde{\theta}_k)^2 \sigma^2 + \sum_{k=0}^{h-1} \theta_k^2 \sigma^2. \tag{6}$$

Thus, we find that $\Delta\mathrm{MSE}_h \equiv \mathrm{MSE}_{h+1} - \mathrm{MSE}_h$ is

$$\Delta\mathrm{MSE}_h = \theta_h^2 \sigma^2 - (\theta_h - \tilde{\theta}_h)^2 \sigma^2, \tag{7}$$

which gives the improvement in forecast content with the new information. The first element on the RHS represents the maximum possible improvement in the quality of forecasts if the information is used efficiently, but the second component represents the mistakes in the utilization of the new information. If the usage of the most recent information $\tilde{\theta}_h$ differs from its optimal value θ_h, the gain from the utilization of new information will decrease and result in excess variability in the forecasts given by the second term. In the special case where $\tilde{\theta}_h = \theta_h$, Equation (7) is equivalent to Equation (3). In this case, $\Delta\mathrm{MSE}_h$ will measure the content of new information in the actual process, which is simply $\theta_h^2\sigma^2$. Note, however, that a nonnegative MSE differential is compatible with the situation where $\tilde{\theta}_h \neq \theta_h$ for a wide range of parameter values. This underscores the point that the bounds derived by PT are implied by forecast rationality, and hence are not necessary conditions. In other words, if these tests reject the null, we have evidence against forecast rationality, but if the tests do not reject, we cannot say we have evidence in favor of rationality. The issue is whether extant forecast efficiency tests, like those due to Nordhaus (1987) or Davies and Lahiri (1999), would detect forecast irrationality under the latter scenario.

While the use of $\Delta\mathrm{MSE}_h$ provides an estimate of the improvement in forecasting performance at horizon h in an ex-post sense, a similar measure can be constructed based solely on forecasts without using the actual data on the target variable—a point emphasized by PT. Notice that, based on Equation (1), the optimal forecast revision $r_{t,h} \equiv f_{t,h} - f_{t,h+1}$ is nothing but $r_{t,h} = \theta_h \varepsilon_{t-h}$. In the suboptimal case of Equation (4), we have the forecast revision process $r_{t,h} = \tilde{\theta}_h \varepsilon_{t-h}$. Calculating the mean squared revisions (MSRs) at horizon h and taking the probability limit, we get $\mathrm{MSR}_h = p\lim_T \frac{1}{T}\sum_{t=1}^{T} r_{t,h}^2 = \tilde{\theta}_h^2\sigma^2$, which provides a measure for the reaction of the forecasters to news. But since forecasters react to news based on their perceptions of the importance of the news, this measure can be seen as the content of the new information as perceived by the forecasters in real time. Note the clear difference between $\Delta\mathrm{MSE}_h$ and MSR_h. While the former is driven by the forecast errors, the latter has nothing to do with the actual process or the outcomes. But both of the measures should give the same values if the survey forecasts are optimal. This result was originally pointed out by Isiklar and Lahiri (2007).

The difference between MSR_h and $\Delta\mathrm{MSE}_h$ may provide important behavioral characteristics of the forecasters such as over or underreaction to news at a specific forecast horizon. MSR_h can be seen as a measure of how forecasters interpret the importance of news at a specific horizon, and $\Delta\mathrm{MSE}_h$ can be seen as the “prize” they get as a result of revising their forecasts. Suppose that forecasters make large revisions at horizon h, but the performance of the forecasts does not improve much at that horizon, then one may conjecture that the forecasters react excessively to the news. To see this more clearly, simple algebra yields $\mathrm{MSR}_h - \Delta\mathrm{MSE}_h = 2(\tilde{\theta}_h^2 - \theta_h\tilde{\theta}_h)\sigma^2$, which is positive

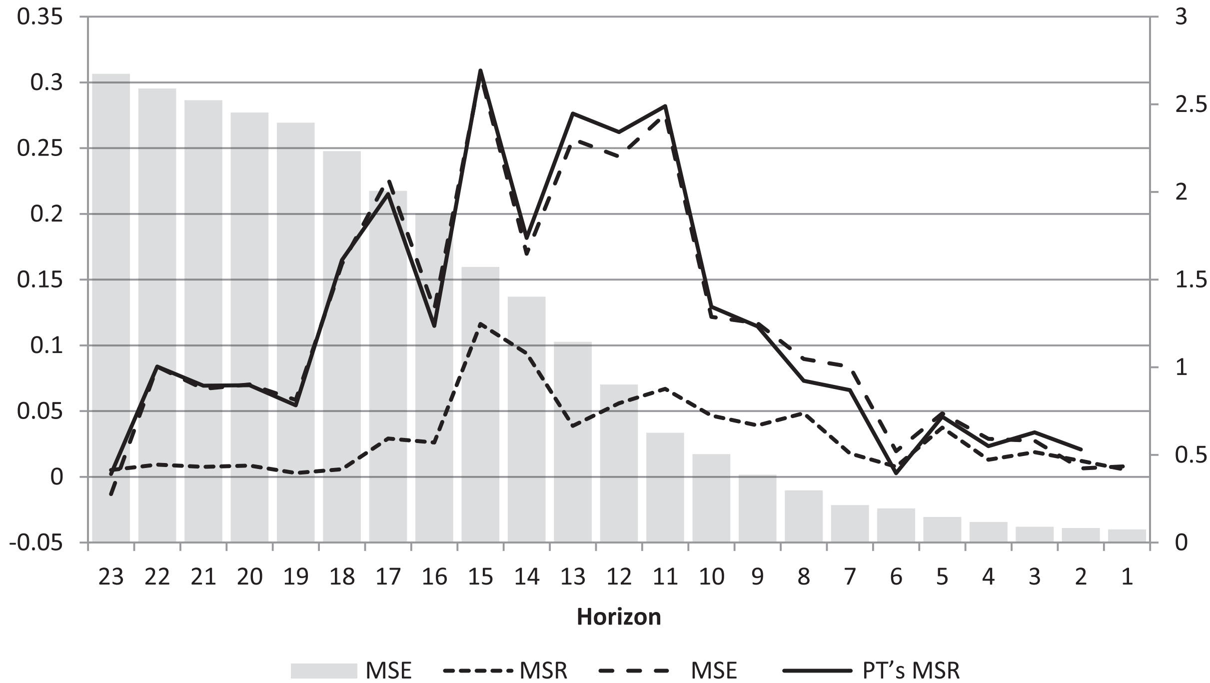

Figure 2. MSE, MSF, and their differentials, Blue Chip Consensus Forecasts 1986–2009.

when $\tilde{\theta}_h^2 > \theta_h\tilde{\theta}_h$. This is the same as the condition $|\tilde{\theta}_h| > |\theta_h|$. But $|\tilde{\theta}_h| > |\theta_h|$ is equivalent to overreaction to the news when the horizon is h. Thus, forecast optimality can be tested by the equality between MSR_h and $\Delta\mathrm{MSE}_h$. This concept is the basis of the amended Nordhaus-type rationality test suggested by Lahiri and Sheng (2008) in which forecast error is regressed on the latest forecast revision for testing its significance. Note that PT's bounds test on the MSRs is defined slightly differently as the difference between $f_{t,1} - f_{t,m}$ and $f_{t,1} - f_{t,m-1}$ where m is an intermediate horizon.

3. ADDITIONAL SURVEY EVIDENCE

In order to visualize the differences in MSR_h as we defined above, $\Delta\mathrm{MSE}_h$ and PT's MSR, we plotted these values in Figure 2, calculated from the Blue Chip consensus forecasts over 1986–2009. The actual values are the first available real-time data obtained from the Philadelphia Fed's real-time database. First note that PT's MSR and $\Delta\mathrm{MSE}_h$ are almost identical at most horizons because when the shortest horizon forecast is very close to the actual value, as is actually the case, the two measures will be very similar. Second, both are nonnegative at all horizons suggesting forecast rationality by the PT criteria. However, we find a substantial wedge between MSR_h and $\Delta\mathrm{MSE}_h$ particularly in the middle horizons that suggests substantial underreaction to news, and hence inefficiency. Thus, PT's MSE and MSR differential bounds conditions are not stringent enough to detect inefficiency in the Blue Chip consensus forecasts. This is consistent with the evidence they report. Note, however, with the same Blue Chip data series over 1977–2009, the Nordhaus test readily rejects rationality over multiple horizons with adjusted R^2 in excess of 0.20 and the coefficient on the lagged forecast revision around 0.58. This result is valid over different sample periods 1977–2009 and 1986–2009, over all horizons and also over horizons between 16 and 6. Note that even though PT did not consider the Nordhaus test in their experiments, their bounds test that the variance of the forecast revision should not exceed twice the covariance between the forecast revision and the actual value is effectively the Nordhaus test in disguise because, given that the longer horizon forecasts have larger MSEs than shorter horizon forecasts, this particular PT condition is derived using the Nordhaus condition that forecast errors should be uncorrelated with forecast revisions under forecast efficiency (see their proof of Corollary 4 in the appendix).

We also experimented with the extended MZ regression using consensus Blue Chip real GDP forecasts from 1977 to 2009. Data on horizons 16 through 7 are available throughout the Blue Chip sample. PT's univariate optimal revision regression generalizing the MZ regression rejected the null that the intercept is zero and that the coefficients of the horizon 16 forecast and the series of intermediate forecast revisions are one with the p-value of 0.07. But all individual MZ regressions accepted the unbiasedness hypothesis with p-values in excess of 0.5. This result is very similar to what PT found with Greenbook forecasts on real GDP. However, the high multicollinearity between successive revisions tends to make this regression highly unstable, particularly when forecasts on a large number of horizons are available. Thus, one should be careful while using this test—the conclusions using this test may depend on the horizons included in the extended regression.

I also used another rich survey panel dataset—the U.S. Survey of Professional forecasters (SPF)—over 1968Q4–2011Q1 using three primitive forecasts of individuals (ID nos 40, 65, and 85) each having more than 100 quarters of participation, and also all forecasters who participated at least 10 times yielding a total of 425 forecasters in the "all" group. I used real GDP forecasts for six available horizons—beginning with the current quarter. Various forecast statistics are reported in Table 1. The actual

Table 1. Real GDP forecast error statistics for SPF data

	Forecaster ID no. 40			Forecaster ID no. 65			Forecaster ID no. 85			All		
Horizon quarter	MSE	MSR	ΔMSE	MSE	MSR	ΔMSE	MSE	MSR	ΔMSE	MSE	MSR	ΔMSE
1	7.295	2.997	0.788	4.447	5.645	4.62	4.572	2.283	2.66	7.083	7.203	3.793
2	8.083	1.765	3.442	9.067	2.962	2.066	7.232	1.789	−0.789	10.876	6.650	2.444
3	11.525	0.766	1.147	11.133	2.927	0.474	6.443	1.326	−0.148	13.320	6.696	3.543
4	12.672	1.495	−4.129	11.607	4.934	3.69	6.295	2.142	5.057	16.863	5.601	−0.065
5	8.543	-	-	15.297	-	-	11.352	-	-	16.798	-	-

GDP values are again the real-time figures released one month after the end of the quarter. Here, we find very few negative ΔMSE_h or MSR values that would suggest inefficiency. Only the MSE differential for forecaster 40 between quarter 5 and quarter 4 is substantially negative, suggesting forecast suboptimality. However, this evidence of inconsistency can be a result of the arrival of the current year's real GDP value for predicting the first quarter GDP growth for the next year. More generally, relevant information regarding different target values may arrive at different times in a nonmonotone manner, and as a result, the relative forecast accuracy over horizons may not be smooth. PT's approach of pooling all horizons together to test forecast rationality can mask this important horizon-specific heterogeneity in forecast efficiency. In other words, forecasts may be efficient at certain horizons but not at others.

In Figure 3, we have plotted total sum of squares of forecast revisions (defined as $S_h^t = \sum_{i=1}^{N_h} \sum_{t=1}^{t_h^i} \frac{(r_{t,h}^i - \overline{\overline{r}}_h)^2}{\sum_{i=1}^{N_h} T_h^i}$, where i refers to the ith forecaster, $r_{t,h}^i = f_{t,h}^i - f_{t,h+1}^i$, $\overline{\overline{r}}_h = \frac{1}{N_h} \sum_{i=1}^{N_h} \frac{1}{T_h^i} \sum_{t=1}^{T_h^i} r_{t,h}^i$, and N_h and T_h^i denote the available observations) at horizons from 23 to 1 to illustrate that the maximum amount of revisions take place in the middle horizons, and interestingly, the maximum amount of underreaction to news takes place at these horizons too. See Lahiri and Sheng (2010) for additional evidence on this point. These results mean that the efforts put forward by forecasters to produce serious forecasts vary by horizons, and the process begins seriously at around 16-month horizon. Thus, forecasting efforts and the resulting efficiency are also conditioned by the institutions' requirements under which the forecasters operate. To truly understand the forecasting inefficiencies, the demand side of the forecasting market should also be considered, in addition to the schedule of official data announcements.

There are a few more such negative (though small) MSE differentials in Table 1. PT also found a similar result with respect to Greenbook real GDP forecasts; see also Clements et al. (2007). We find that the MSE differentials and MSRs are quite different particularly in the middle horizons and the latter tend to underestimate the former, suggesting underreaction to new information. Following PT, we also calculated MSR differentials between $f_{t,2} - f_{t,4}$ and $f_{t,2} - f_{t,3}$, and $f_{t,3} - f_{t,5}$ and $f_{t,3} - f_{t,4}$; and between $f_{t,4} - f_{t,6}$ and $f_{t,4} - f_{t,5}$ for the three long-standing forecasters and also for the "all' group using disaggregate data. In none of the cases did we find any evidence of negative MSR differentials and thus we fail to detect any indication of irrationality based on this MSR criterion. However, the Nordhaus test and the regressions of forecast errors on forecast revisions, *a la* Lahiri and Sheng (2008, 2010), readily detected deviations from rationality in most cases.

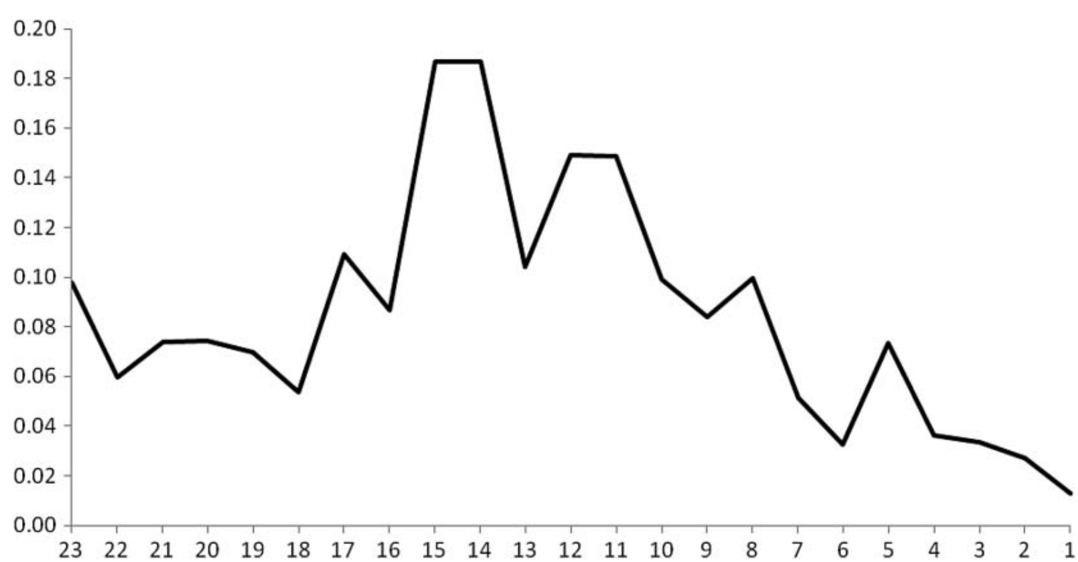

Figure 3. Total sum of squares in GDP forecast revisions during 1986–2009: Blue Chip Surveys.

4. CONCLUSION

I find the ideas put forward in this article to rigorously test the optimality of forecasts across horizons interesting and the efforts to implement the tests quite commendable. Despite their mathematical elegance, these derived bounds are implied by forecast rationality, and hence are not necessary conditions. Thus, if the tests do not reject the null, we cannot say we have evidence in favor of rationality. For instance, even though the forecast variances and MSEs are observed to be weakly decreasing functions of the forecast horizon, the underlying forecasts can be easily be still inefficient. By using Blue Chip and SPF survey forecasts, we found, like PT, that some of the bounds tests proposed in the paper are not very powerful to detect suboptimality in instances where the extant Nordhaus test readily identifies it. However, their new extended MZ test based on a regression of the target variable on the long-horizon forecast and the sequence of interim forecast revisions works well, provided the multicollinearity problem does not become serious. We have argued that by testing the equality of MSE differentials with mean square forecast revisions, one can also examine forecast rationality over multiple horizons. In order to truly understand the pathways through which forecasts fail to satisfy forecast optimality, we have also to consider the demand side of the forecasting market and the schedule of official data announcements. For instance, there is evidence that forecasters record maximum suboptimality at horizons where they also make maximum forecast revisions. The observed suboptimality of Greenbook forecasts that PT found cannot be understood unless the institutional requirements of such forecast are appreciated. Nevertheless, the importance of testing the joint implications of forecast rationality across multiple horizons when such information is available as proposed by the authors must be appreciated.

ACKNOWLEDGMENTS

The author thanks Antony Davies, Gultikin Isiklar, Huaming Peng, Xuguang Sheng, and Yongchen Zhao for many useful discussions and research assistance, and to the co-editors for making a number of helpful comments.

[Received June 2011. Revised July 2011.]

REFERENCES

Clements, C. P., Joutz, F., and Stekler, H. (2007), "An Evaluation of Forecasts of the Federal Reserve: A Pooled Approach," *Journal of Applied Econometrics*, 22, 121–136. [24]

Davies, A., and Lahiri, K. (1999), "Re-examining the Rational Expectations Hypothesis Using Panel Data on Multi-period Forecasts," in *Analysis of Panels and Limited Dependent Variables*, eds. C. Hsiao, K. Lahiri, L. F. Lee, and H. Pesaran, Cambridge, U.K.: Cambridge University Press, pp. 226–254. [22]

Isiklar, G., and Lahiri, K. (2007), "How Far Ahead can We Forecast? Evidence from Cross-country Surveys," *International Journal of Forecasting*, 23, 167–187. [21,22]

Lahiri, K., and Sheng, X. (2008), "Evolution of Forecast Disagreement in a Bayesian Learning Model," *Journal of Econometrics*, 144, 325–340. [23,24]

Lahiri, K., and Sheng, X. (2010), "Learning and Heterogeneity in GDP and Inflation Forecasts," *International Journal of Forecasting*, 26, 265–292. [24]

Mankiw, N. G., and Shapiro, M. D. (1986), "News or Noise? An Analysis of GNP Revisions," *Survey of Current Business*, 66, 20–25. [21]

Nordhaus, W. (1987), Forecasting Efficiency; Concepts and Applications, *Review of Economics and Statistics*, 69, 667–674. [22]

Comment

Barbara Rossi

Department of Economics, 204 Social Science Building, Duke University, Durham, NC 27708, ICREA, and University Pompeu (*brossi@econ.duke.edu*)

Patton and Timmermann (2011) propose new and creative forecast rationality tests based on multi-horizon restrictions. The novelty is to consider the implications of forecast rationality jointly across the horizons. They focus on testing implications of forecast rationality such as the fact that the mean squared forecast error should be increasing with the forecast horizon (Diebold 2001; Patton and Timmermann 2007) and that the mean squared forecast should be decreasing with the horizon. They also consider new regression tests of forecast rationality that use the complete set of forecasts across all horizons in a univariate regression, which they refer to as the "optimal revision regression" tests. One of the advantages of the proposed procedures is that they do not require researchers to observe the target variable, which sometimes is not clearly available. In fact, Patton and Timmermann (2011) show that both their inequality results as well as the "optimal revision regression" test hold when the short horizon forecast is used in place of the target variable. Their work is an excellent contribution to the literature.

The main objective of this comment is to check the robustness of forecast rationality tests to the presence of instabilities. The existence of instabilities in the relative forecasting performance of competing models is well known [see Giacomini and Rossi (2010) and Rossi and Sekhposyan (2010), among others; Rossi (2011) provide a survey of the existing literature on forecasting in unstable environments]. First, we show heuristic empirical evidence of time variation in the rolling estimates of the coefficients of forecast rationality regressions. We then use fluctuation rationality tests, proposed by Rossi and Sekhposyan (2011), to test for forecast rationality, while, at the same time, being robust to instabilities. We also consider a version of Patton and Timmermann's (2010) optimal revision

DOI: 10.1080/07350015.2012.634343