Research Program on Forecasting

What Can We Learn From Revisions to the Greenbook Forecasts?

Jeff Messina
The George Washington University
jeffreym@gwmail.gwu.edu

Tara M. Sinclair
The George Washington University
tsinc@gwu.edu

Herman O. Stekler
The George Washington University
hstekler@gwu.edu

RPF Working Paper No. 2014-003
http://www.gwu.edu/~forcpgm/2014-003.pdf

June 21, 2014

RESEARCH PROGRAM ON FORECASTING
Center of Economic Research
Department of Economics
The George Washington University
Washington, DC 20052
http://www.gwu.edu/~forcpgm

What Can We Learn From Revisions to the Greenbook Forecasts?

Jeff Messina, Tara M. Sinclair, and Herman Stekler

Department of Economics

The George Washington University

Keywords: Federal Reserve; Forecast Evaluation; Forecast Revisions

JEL Codes: C5, E2, E3

Abstract

Although there have been many evaluations of the Fed Greenbook forecasts, we analyze them in a different dimension. We examine the revisions of these forecasts in the context of fixed event predictions to determine how new information is incorporated in the forecasting process. This analysis permits us to determine whether there was an underutilization of information. There is no evidence of forecast smoothing, but rather that the revisions were sometimes in the wrong direction.

What Can We Learn from Revisions to the Greenbook Forecasts?

Over the years, the economic forecasts of the Federal Reserve staff, reported in the Greenbook, have been evaluated extensively. These evaluations have primarily focused either on forecast accuracy or sought to determine whether the forecasts were rational. Previous studies, however, have given little consideration to the informational content of the forecast revisions.

Several statistical techniques have been used to evaluate these Fed's forecasts. Most analyses used statistical tests that focused either on a comparison of the forecasts with the outcomes (Mincer-Zarnowitz, 1969) or examined the characteristics of the errors (Holden-Peel, 1990). There is a third statistical test for evaluating univariate forecasts that until recently had not been used as extensively to test the efficiency of forecasts. That technique developed by Nordhaus (1987) examines the sequential revisions of the forecasts of a particular event rather than the forecasts themselves or their errors. The results of that test also permit one to determine how well new information is used in revising forecasts. Thus the main contribution of this paper is to examine the role of forecast revisions within the Fed's forecasting process and to determine how the Fed's forecasting staff weighs new information. This is a subject that has not received much attention in the previous analyses of the Greenbook predictions.

The following section presents a summary of previous findings about the Greenbook forecasts. We then present the alternative ways of measuring bias and discuss our analysis of the revisions of the forecasts, the interaction of those revisions with a variable that represents the state of the economy and the implications of those results as it relates to the use of new

information. The concluding sections compare our results with those obtained from other forecasters and summarize our results pertaining to the use of new information.

I. Previous Evaluations of Fed Forecasts

The Greenbook forecasts made by the Federal Reserve staff have been extensively evaluated. The forecasts are made several times each quarter with horizons of zero to eight quarters. Some analyses of these forecasts of real GDP growth, inflation and unemployment have shown that they contain systematic errors. Using *qualitative* analyses the studies showed that growth was underestimated (overestimated) when the economy was increasing (decreasing).[1] Similar results were found for the inflation forecasts when inflation was accelerating or decelerating. Finally, the recessions that occurred between 1965 and 2007 were not predicted in advance.

Despite these errors, the overall *quantitative* evidence is that the Greenbook forecasts are as good or better than the predictions made by private sector forecasters. (See Romer and Romer, 2000; Sims, 2002; Gamber and Smith, 2009).[2] Romer and Romer showed that the Greenbook forecasts were always better than the commercial predictions; this was especially true for the inflation forecasts. The later study by Gamber and Smith corroborated this finding, but argued that while the advantage was still significant, it had lessened.[3] Moreover, Faust and Wright (2009) compared the Greenbook forecasts with time series models using real-time data. They concluded: "… we again find that Greenbook is a very good forecast for both output growth and inflation." (p.478).

[1] Fildes and Stekler (2002) found similar results in the forecasts of other individuals or institutions.
[2] Hubert (2011) summarizes the specific results of the studies that evaluated the Fed forecasts. He indicates that the Fed's inflation forecasts were definitely better than those of the private sector, while the two sets of growth forecasts were about the same.
[3] On the other hand, Joutz and Stekler (2000) showed that the Fed forecast errors were only comparable to those made by the private sector economists.

II. Alternative Measures of Bias and Efficiency

A. Previous Analyses

Optimal forecasts should be unbiased and there are two commonly used procedures for testing for bias. One compares the forecasts ($F_{t,h}$) with the outcomes (A_t), following Mincer-Zarnowitz (1969):

$$A_t = a_0 + a_1 F_{t,h} + u_t, \quad (1)$$

where *h* is the horizon of the forecast and the joint null of unbiasedness is that $a_o = 0$ and $a_1 = 1$. The other examines the characteristics of the forecast errors (Holden-Peel, 1990):

$$e_{t,h} = (A_t - F_{t,h}) = a_0 + u_t \quad (2)$$

In this case the null is that $a_0 = 0$.[4]

As mentioned above, there is a third approach for evaluating economic forecasts. It deals with the efficiency of the forecasts and focuses on the revisions of the forecasts rather than on the forecasts themselves or the errors that were made. Nordhaus (1987) provides two propositions about the behavior of forecast revisions if the forecasts are efficient. The first indicates that the forecast *error* at time t-h for the target date, t, is independent of all previous forecast revisions. The second proposition is that the forecast *revision* at any time t is independent of all previous revisions. Thus efficient forecasts required that revisions should be unpredictable and should form a martingale. Although these fixed-target tests could have more power (Nordhaus, p. 669) in testing for efficiency than do the Mincer-Zarnowitz and Holden-

[4]Many of studies showed that the Greenbook predictions of one or more variables at one or more horizons failed to take into account available information or displayed ex post bias. (See for example, Jansen and Kishan, 1996 and 1999; Joutz and Stekler, 2000; Clements, Joutz and Stekler, 2007; Sinclair, Joutz and Stekler, 2010)). The results depended upon the dates of the forecasts that were being analyzed because the systematic errors, associated with the phases of the cycles, tended to offset each other when either Eq. (1) or Eq. (2) was used. (See Sinclair et al., 2010).

Peel statistics, the former have not been used as extensively. While Clements et al. (2007) and Patton and Timmermann, (2012) used this approach in examining the revisions of the Greenbook forecasts, neither focused on the signs of the revisions to determine whether the Fed staff over or under adjusted their forecasts in response to the new information.

B. Our Approach

We will only discuss the first of these two Nordhaus tests because it enables us to make inferences about the use of new information in the forecasting process. This proposition can be expressed by Eq. 3:

$$(A_t - f_{t,h}) = a_0 + a_1 (f_{t,h} - f_{t,h+1}) + u_{t,h}. \quad (3)$$

The null is that a_0 and a_1 are jointly equal to zero. If the coefficients are significant, it is possible to interpret them in the context of the forecasting process. A significant a_0 would indicate that the forecasts are either over or underestimated systematically. Conventionally, a positive and significant a_1 coefficient would indicate that all new information had not been fully incorporated into the forecast of t made at time t-h. A significant negative coefficient would suggest that the forecast had been over adjusted based on the new information.

Recently, the same approach was used to examine the characteristics of the revisions of other sets of forecasts to determine how information was used and whether the forecasts were efficient. (See Coibion and Gorodnichenko 2010, 2012; Lahiri and Sheng 2008, 2010; and Loungani, Stekler and Tamirisa 2013). In many cases, the a_1 coefficients were positive indicating that new information had not been completely incorporated into the revisions. Some

authors, therefore, suggested that the forecasts had been smoothed, while others found significant negative coefficients showing that the forecasts had been over adjusted.[5]

III. Data

We examine the Greenbook current quarter and one to four-quarter-ahead forecasts of four variables; (1) real GDP growth, (2) unemployment, and inflation as measured by (3) the GDP deflator and (4) the CPI. The forecasts for the first three variables cover the period 1975.II – 2008.IV,[6] while the CPI predictions are available only from 1980.IV- 2008.IV. Whenever there are multiple forecasts in each quarter, we use the one closest to the middle of the quarter following the Federal Reserve Bank of Philadelphia's Greenbook data set.[7] The actual NIPA data are the real time estimates published approximately 90 days after the quarter to which they refer.[8] All data with the exception of the unemployment rate are converted into annualized growth rates.

IV. Methodology

Using this common data set, we utilize a number of different statistical procedures. The first step of our analysis is to determine whether the forecasts are biased. We use the Holden-Peel (1990) approach and regress the ex post forecast error on a constant (equation 2) and test whether the constant is statistically significant. The next step in the process is to include the forecast revision in Equation 2. This relationship is captured by equation 3. If last period's ex ante forecast revision can help to explain a portion of this period's ex post forecast error, the new

[5] In any event, Chang et al. (2013) have indicated that it is not always possible to interpret the coefficient in that way.

[6] This time period was chosen because we wanted to be able to compare the current quarter and one to four quarter ahead forecasts on a comparable basis. The Greenbook did not contain four-quarter-ahead forecasts on a consistent basis until 1975. There is a five year lag in publishing the Greenbooks.

[7] The full dataset and documentation are available here: http://www.phil.frb.org/research-and-data/real-time-center//greenbook-data/.

[8] This is the "third release" from the Philadelphia Fed's Real Time Data Set, available here: http://www.phil.frb.org/research-and-data/real-time-center/real-time-data/

information that became available in the prior period was not properly utilized, and the forecast could have been over or under adjusted.

Equations 2 and 3 would have to be modified if there were systematic biases that cancel each other. As noted above, previous research has shown that there are systematic biases that are related to the state of the economy. Consequently, the systematic errors that are related to the phases of the various cycles would cancel each other, and the Holden-Peel test would fail to reject the null when in fact there was a systematic ex post bias.

Consequently, we modify Equations 2 and 3 to introduce a dummy D_t that reflects the state of the economy following Sinclair, Joutz, and Stekler (2010). In the growth and employment equations, the dummy has a value 1 if in any month of a particular quarter the economy was in a National Bureau of Economic Research dated recession and is zero otherwise. The joint null is that all the coefficients are zero. If the dummy coefficients are non-zero, they contain information indicating that the forecasts did not incorporate knowledge about the future state of the economy. Similarly, if the coefficient on the forecast revision is significant this indicates that there is information in that variable that can explain the current error.

$$e_{t,\,h} = a_0 + a_2 D_t + u_{t,\,h} \tag{2a}$$

$$e_{t,\,h} = a_0 + a_1(f_{t,h} - f_{t,\,h+1}) + a_2 D_t + a_3 D_{t*}(f_{t,\,h} - f_{t,\,h+1}) + u_{t,\,h}. \tag{3a}$$

There is no bias if the null that all of the coefficients in Equations 2a and/or 3a are zero is not rejected. If the coefficient on the revision is significantly different from zero, the new information was interpreted inefficiently; either too little or too much was incorporated into the new forecast. We also considered the possibility that the Fed staff weighed information

differently in recessions versus expansions and interacted the revisions and state of the economy terms of the equations.

V. Results

A. Holden-Peel Test

The results in Table 1 show that the current quarter forecasts of real GDP and the unemployment rate reject the null of no bias at the 5% level of significance. Real GDP has a consistently positive bias (although insignificant after the current quarter) while the other three variables have a consistently negative bias (although insignificant for all horizons for the deflator and CPI inflation, but significant for the current quarter and the 1 and 2 quarter ahead forecasts for the unemployment rate).[9]

B. Revisions of the Forecasts

The results presented in Table 2, based on equation 3, show that there is not a single variable at any forecast horizon where there is significant underutilization of information. This type of inefficiency would have required that the coefficient on the revisions variable be positive and significant. Instead most coefficients are insignificant. When those coefficients were significant, they were negative. This was the case for the growth forecasts made between h = 2 and h = 1 and h = 4 and h = 3 as well as for the deflator for the revision of the forecast made between h = 1 and h = 0 and for CPI inflation for the revision of the forecast made between h = 3 and h = 2. In these cases instead of responding efficiently to new information the Fed appears to have put too much weight on new information in revising their forecasts. A smaller revision would actually have lowered the forecast error in these cases. We did not find any significant

[9] The real GDP and deflator biases are in opposite directions. If the forecasters were analyzing nominal GDP data, there is a failure to correctly divide these numbers into the two components.

positive coefficients on the revisions which suggests that any observed biases cannot be explained by underutilization of the available information.

C. State of the Economy

The qualitative analyses of the Fed forecasts have shown the presence of systematic errors. We have just shown that there is no reason to argue that new information is not fully incorporated into the revised forecasts. We next examine the relationship between forecast revisions and the forecasts that were associated with two phases of the economy.

Equation 2a is used to determine whether the inclusion of the state of the economy dummies yields results that are different from those obtained from the traditional tests for bias. The results (Table 3) for the *current* quarter show that that the coefficients associated with the dummy in the GDP and both inflation equations are not significant. This implies that the Fed's nowcasts correctly incorporated information about the state of the economy. In the current quarter unemployment equation, however, the dummy is significant. This suggests that the Fed appears to have underestimated the current quarter unemployment rate in recessions.[10]

However, the results are different for forecasts made for horizons of one to four quarters. Now the coefficients on the dummy variable are all significant for GDP and the unemployment rate. In the GDP equations, the coefficient on the NBER dummy is negative indicating that the Fed overestimated real growth during recessionary periods. In the unemployment equation the dummy is positive showing that unemployment was underestimated, a result consistent with the GDP projections. Furthermore, in several cases the constants were significant. Thus there were biases even after accounting for the failure to incorporate information about the state of the

[10] This finding appears dependent on the end of the sample for the 2008 period. If we exclude this recession then the coefficient on the revision is insignificant.

economy. For the inflation equations, however, the NBER dummy and the constants were never significant.[11]

D. Revisions and State of the Economy

Do our findings about the forecast revisions still hold when we account for the errors made by not predicting the state of the economy? To test this proposition, we estimate Equation 3a where we interact the state of the economy dummy and the forecast revision. A significant coefficient on the interaction term would show that revisions were made a different way if there were a recession rather than a period of growth. Similarly, the interaction term in the inflation equations would indicate whether the revisions were made differently during periods of acceleration and deceleration of inflation.

We continue to find that there are no significant positive coefficients associated with the revisions variable, even when we include a dummy variable for the state of the economy and interact the dummy and the forecast revision. (Tables 4a and 4b). There continue to be significant negative coefficients associated with the growth projections, indicating that the revisions were in the wrong direction. The results in Tables 4a and 4b indicate that only for the deflator were coefficients associated with the interaction terms significant.

Thus, the impact of revisions on the Fed's forecast errors did not differ between the states of the economy except perhaps for the case of the deflator. This result at least suggests that the Fed staff did not place different weights on errors made during recessions and periods of growth, i.e. there was no asymmetric loss function in this dimension, perhaps because they did not know the state of the economy at the time they were making the forecasts.

[11] This is consistent with the findings of Sinclair, Joutz and Stekler (2010) who also found no significant relationship between inflation forecast errors and the NBER dummy. Instead, they found a relationship when they used a measure of the inflation cycle obtained from ECRI to recognize and incorporate information about the accelerations/ decelerations of inflation.

VI. Interpreting the Results

Unlike some previous studies of other forecasts that found that information was underutilized, (e.g. Coibion and Gorodnichenko 2010, 2012; Loungani et al. 2014), we did not obtain this result for the Fed's Greenbook forecasts. Rather, we found no evidence of underutilization and also some over adjustments to new information. We, thus, sought to compare our results with those obtained from other data sets. We first examined published studies that had also examined forecast revisions. We then conducted an identical analysis of the predictions in the Survey of Professional Forecasters.

Using *Consensus Economics* data, Lahiri and Sheng (2008) had evaluated the consensus real GDP forecasts made for the G-7 countries. They showed that the results varied from country to country and from horizon to horizon. In many cases, the information available at long horizons was overweighted, i.e. the coefficients on the revisions were negative and statistically significant. At middle length horizons, the opposite results were obtained, i.e. positive and significant.

Dovern and Weisser (2011) also examine the G-7 forecasts obtained from *Consensus Economics*, but they utilize the individual predictions rather than the mean forecast. Most of the time, they do not reject the hypothesis that the forecasts of inflation, industrial production and private consumption are weakly efficient. When this hypothesis is rejected, it is because the forecasters have overreacted to the new information. Only the real GDP predictions display substantial evidence that the individual forecasters underreacted to the new information or smoothed the forecasts. Since there is no uniformity in the results, one can conclude that the way that forecasters use new information in revising their predictions differs from variable to variable.

The study by Dovern, Fritsche, Loungani and Tamirisa (forthcoming) yields an additional insight. They also used the *Consensus Economics* forecasts of GDP but expanded the analysis in two dimensions. First, they examined the revisions of the GDP forecasts for both advanced and emerging economies. Second and most important they compared the individual forecasts with the mean forecasts for each country. They concluded that the level of rigidity in the average forecasts was similar to that which Coibion and Gorodnichenko had found.[12] Moreover, there also was rigidity in the individual predictions but it was less than in the mean predictions. They concluded that the process of averaging the individual forecasts induces additional stickiness or smoothing.[13]

Finally, we want to compare the results for the SPF forecasts with those that we obtained from the Greenbook projections. We used the same time period and identical procedures in this comparison. Our results (See Appendix Tables A1 and A2) indicate the forecasters smoothed the forecasts for some variables and overreacted to new information when predicting other variables. The unemployment forecasts were smoothed while the inflation predictions show evidence of over adjustment. There are even some small differences in the results if the mean forecast is used rather than the median.[14]

These results lead us to conclude that there is no unique single overall explanation, model, or theory of the way that forecasters utilize information.[15] Some may smooth their forecasts; others may over adjust to new information. Moreover, each individual's adjustment

[12] However, Dovern et al. did not present the number of times that information was underutilized or the number of times that forecasters overreacted.
[13] Capistran and Lopez-Moctezuma (2014) also found that aggregating increased the tendency towards smoothing.
[14] For example at the two and three quarter horizons, the real GDP coefficients are not significant in one case but are significant at the 10% level in the other instance. A similar result can be observed in the deflator revisions at the one quarter horizon.
[15] This result is in accord with the findings of Andrade and Bihan (2013) who argued that it was not possible to build or estimate a model that was consistent with the errors and disagreements that were observed in the data.

process may differ from variable to variable. For that reason it is important to examine each individual's (organization's) ability to process and utilize new information.

VII. Conclusions

We have two major conclusions. Our results quite clearly indicate that the Federal Reserve's Greenbook forecasts display some biases, but that the underutilization of information or forecast smoothing could not explain these inefficiencies. In most cases, the Fed staff utilized new information efficiently to update the forecasts of all variables not only for the current quarter but also for distant periods. There were no significant positive coefficients, a result that would have indicated that the forecasts had been smoothed. The departures from efficiency occurred as a result of over reacting to new information.

Acknowledgements

The authors gratefully acknowledge the support of Luther Rice Undergraduate Research Fellowship for this project.

Table 1
Holden-Peel Test for Bias
current quarter and one to four quarters ahead

Horizon / Variable	$h = 0$	$h = 1$	$h = 2$	$h = 3$	$h = 4$
GDP	0.44** (0.19)	0.30 (0.28)	0.11 (0.30)	0.05 (0.30)	0.02 (0.31)
Unemployment	-0.05*** (0.01)	-0.09** (0.04)	-0.12* (0.06)	-0.13 (0.09)	-0.14 (0.011)
Deflator	-0.13 (0.09)	-0.08 (0.13)	-0.07 (0.14)	-0.06 (0.16)	-0.06 (0.18)
CPI	-0.10 (0.11)	-0.16 (0.25)	-0.12 (0.26)	-0.08 (0.25)	-0.18 (0.26)

*, **, ***: Significant at the 10%, 5%, and 1% level respectively. Newey-West standard errors reported for one to four quarters ahead.

Table 2

Nordhaus test with the Fed's forecast error as the dependent variable

Horizon	$h = 0$		$h = 1$		$h = 2$		$h = 3$	
Variable	***Constant***	***Revision***	***Constant***	***Revision***	***Constant***	***Revision***	***Constant***	***Revision***
GDP	0.47** (0.18)	0.15 (0.14)	0.20 (0.28)	-0.53** (0.26)	0.13 (0.30)	0.42 (0.32)	0.03 (0.31)	-0.83** (0.33)
Unemployment	-0.05** (0.02)	-0.02 (0.05)	-0.10** (0.04)	-0.01 (0.10)	-0.12* (0.06)	0.05 (0.15)	-0.13 (0.09)	0.14 (0.14)
Deflator	-0.11 (0.10)	-0.35** (0.14)	-0.08 (0.13)	0.01 (0.16)	-0.07 (0.15)	-0.40 (0.33)	-0.06 (0.16)	0.09 (0.37)
CPI	-0.09 (0.11)	0.07 (0.13)	-0.16 (0.25)	-0.05 (0.30)	-0.09 (0.25)	-0.79** (0.34)	-0.12 (0.24)	-0.24 (0.61)

*, **, ***: Significant at the 10%, 5%, and 1% level respectively. Newey-West standard errors reported.

Table 3

Holden-Peel Test for Bias including NBER recession dummy
current quarter and one to four quarters ahead

Horizon	***h* = 0**		***h* = 1**		***h* = 2**		***h* = 3**		***h* = 4**	
Variable	***Cons.***	***NBER***	***Cons.***	***NBER***	***Cons.***	***NBER***	***Cons.***	***NBER***	***Cons.***	***NBER***
GDP	0.46** (0.21)	-0.10 (0.53)	0.58* (0.29)	-1.80*** (0.55)	0.53* (0.29)	-2.73*** (0.51)	0.55* (0.30)	-3.21*** (0.48)	0.55* (-3.40)	-3.40*** (0.49)
Unemp.	-0.06*** (0.02)	0.09** (0.04)	-0.15*** ().04)	0.37*** (0.09)	-0.20*** (0.06)	0.52*** (0.16)	-0.23*** (0.08)	0.68*** (0.20)	-0.26** (0.10)	0.79*** (0.21)
Deflator	-0.10 (0.10)	-0.18 (0.25)	-0.04 ().14)	-0.27 (0.31)	-0.03 (0.16)	-0.27 (0.35)	-0.03 (0.17)	-0.23 (0.43)	-0.02 (0.18)	-0.29 (0.51)
CPI	-0.04 (0.12)	-0.37 (0.31)	-0.07 (0.21)	-0.60 (0.81)	-0.05 (0.23)	-0.42 (0.88)	-0.03 (0.24)	-0.37 (0.81)	-0.10 (0.25)	-0.45 (0.82)

*, **, ***: Significant at the 10%, 5%, and 1% level respectively. Newey-West standard errors reported for one to four quarters ahead.

Table 4a

Nordhaus test with the forecast error as the dependent variable including dummy and interaction for t and t+1

Horizon	***h* = 0**				***h* = 1**			
Variable	***Constant***	***Revision***	***NBER***	***NBER* Revision**	***Constant***	***Revision***	***NBER***	***NBER* Revision**
GDP	0.44** (0.18)	0.18 ().18)	0.14 ().83)	-0.04 (0.48)	0.54* (0.29)	-0.72*** ().24)	-3.05*** (0.88)	-0.61 (0.95)
Unemployment	-0.06*** (0.02)	-0.03 ().05)	0.12** (0.05)	-0.12 (0.12)	-0.15*** ().04)	-0.05 (0.08)	0.38*** (0.09)	-0.08 (0.13)
Deflator	-0.09 (0.10)	-0.14 (0.17)	-0.20 (0.32)	-0.58** (0.23)	-0.04 (0.14)	0.04 (0.16)	-0.27 (0.32)	-0.26 (0.52)
CPI	-0.04 (0.11)	-0.02 (0.12)	-0.33 (0.29)	0.18 (0.27)	-0.07 (0.22)	-0.01 (0.32)	-0.59 (0.81)	-0.01 (0.75)

Table 4b

Nordhaus test with the forecast error as the dependent variable including dummy and interaction for t+2 and t+3

Horizon	***h* = 3**				***h* = 4**			
Variable	***Constant***	***Revision***	***NBER***	***NBER* Revision***	***Constant***	***Revision***	***NBER***	***NBER* Revision***
GDP	0.53** (0.25)	0.38 (0.31)	-2.92*** (0.68)	-0.81 (0.64)	0.55* (0.30)	-0.95** (0.37)	-3.44*** (0.52)	-0.27 (1.03)
Unemployment	-0.20*** (0.06)	0.06 ().12)	0.54*** ().14)	-0.30 (0.26)	-0.23*** (0.08)	0.09 (0.16)	0.68*** (0.18)	-0.15 (0.32)
Deflator	-0.03 (0.17)	-0.56 (0.35)	-0.28 (0.34)	0.84* (0.44)	-0.02 (0.17)	-0.24 (0.32)	-0.76 (0.45)	2.24*** (0.44)
CPI	-0.04 (0.24)	-0.60 (0.35)	-0.34 (0.80)	-0.68 (0.81)	-0.03 (0.23)	-0.04 (0.59)	-0.53 (0.84)	-0.96 (1.78)

Table 5
RMSE By Horizon

Horizon / **Variable**	***h* = 0**	***h* = 1**	***h* = 2**	***h* = 3**	***h* = 4**
GDP	2.31	2.93	2.89	3.10	3.05
Unemployment	0.17	0.36	0.53	0.67	0.81
Deflator	1.00	1.10	1.22	1.24	1.35
CPI	1.21	2.07	2.15	2.11	2.13

Appendix

Table A1

Nordhaus test with the median SPF forecast error as the dependent variable

Horizon	***h* = 0**		***h* = 1**		***h* = 2**		***h* = 3**	
Variable	***Constant***	***Revision***	***Constant***	***Revision***	***Constant***	***Revision***	***Constant***	***Revision***
GDP	0.61*** (0.19)	0.28* (0.15)	0.18 (0.24)	-0.27 (0.25)	0.12 (0.25)	0.35 (0.29)	-0.32 (0.25)	-0.49 (0.33)
Unemployment	0.02 (0.03)	0.39** (0.17)	-0.03 (0.03)	0.44*** (0.12)	-0.08* (0.04)	0.46*** (0.14)	-0.09 (0.06)	0.53*** (0.17)
Deflator	-0.14 (0.09)	0.12 (0.15	-0.23** (0.11)	-0.01 (0.21)	-0.28** (0.11)	0.39 (0.27)	-0.36*** (0.12)	0.26 (0.20)
CPI	-0.18 (0.22)	-0.92*** (0.20)	-0.22 (0.20)	-1.05*** (0.21)	-0.22 (0.21)	-0.35 (0.51)	-0.18 (0.21)	0.74 (0.68)

*, **, ***: Significant at the 10%, 5%, and 1% level respectively. Newey-West standard errors reported. Note that the CPI sample in the SPF is 1982Q2-2008Q4.

Table A2

Nordhaus test with the mean SPF forecast error as the dependent variable

Horizon	***h* = 0**		***h* = 1**		***h* = 2**		***h* = 3**	
Variable	***Constant***	***Revision***	***Constant***	***Revision***	***Constant***	***Revision***	***Constant***	***Revision***
GDP	0.63*** (0.19)	0.35* (0.20)	0.20 (0.25)	-0.39 (0.51)	0.19 (0.28)	0.66* (0.34)	-0.20 (0.31)	-0.69* (0.36)
Unemployment	0.03 (0.04)	0.47** (0.19)	-0.04 (0.05)	0.43*** (0.15)	-0.08 (0.06)	0.51** (0.21)	-0.08 (0.08)	0.63*** (0.23)
Deflator	-0.15* (0.08)	0.13 (0.11)	-0.27* (0.14)	-0.31* (0.18)	-0.32** (0.16)	-0.01 (0.39)	-0.36** (0.18)	0.50 (0.35)
CPI	-0.15 (0.17)	-0.89*** (0.23)	-0.18 (0.22)	-0.95*** (0.28)	-0.20 (0.23)	-0.49 (0.70)	-0.19 (0.22)	0.49 (0.66)

*, **, ***: Significant at the 10%, 5%, and 1% level respectively. Newey-West standard errors reported. Note that the CPI sample in the SPF is 1982Q2-2008Q4.

References

Andrade, P. and H.L. Bihan, (2013). Inattentive professional forecasters, *Journal of Monetary Economics*, 60, 967-982.

Capistran, C. and G. Lopez-Moctezuma, (2014). Forecast revisions of Mexican inflation and GDP growth, *International Journal of Forecasting*, 30(2), 177-191.

Chang, C-L., B. de Bruijn, P. H. Franses and M. McAleer, (2013). Analyzing fixed-event forecast revisions, *International Journal of Forecasting*, 29(4), 622-627.

Clements, M. P., F. Joutz, and H.O. Stekler, (2007). An evaluation of the forecasts of the Federal Reserve: A pooled approach, *Journal of Applied Econometrics*, 22(1), 121-136.

Coibion, Olivier (2014). "Comments on Dovern, Fritsche, Loungani and Tamirisa (forthcoming)" *International Journal of Forecasting*, forthcoming.

Coibion, Olivier and Yuriy Gorodnichenko, (2010). "Information Rigidity and the Expectations Formation Process: A Simple Framework and New Facts," NBER Working Paper No. 16537.

Coibion, Olivier and Yuriy Gorodnichenko, (2012). "What Can Survey Forecasts Tell Us About Informational Rigidities?" *Journal of Political Economy* 120(1), 116-159.

Dovern, J. and J.Weisser, (2011). Accuracy, unbiasedness and efficiency of professional macroeconomic forecasts: An empirical comparison for the G7, *International Journal of Forecasting*, 27, 452-465.

Dovern, J., U. Fritsche, P. Loungani, and N. Tamirisa, (forthcoming). Informational rigidities: Comparing average and individual forecasts from a large international panel, *International Journal of Forecasting.*

Faust, J. and J. H. Wright, (2009). Comparing Greenbook and reduced form forecasts using a large realtime dataset, *Journal of Business and Economic Statistics*, 27(4),468-479.

Fildes, R. and H. Stekler, (2002). The state of macroeconomic forecasting, *Journal of Macroeconomics*, 24, 435-468.

Gamber, E. W. and J.K. Smith, (2009). Are the Fed's inflation forecasts still superior to the private sector's*? Journal of Macroeconomics*, 31(2), 240-251.

Holden, K. and D.A. Peel, (1990). On testing for unbiasedness and efficiency of forecasts, *Manchester School*, 48, 120–127.

Hubert, P., (2011). Revisiting the Greenbook's relative forecasting performance, OFCE-SciencesPo, mimeo.

Jansen, D. W. and R.P. Kishan, (1996). An evaluation of Federal Reserve forecasting, *Journal of Macroeconomics*, 18(1), 89-109.

Jansen, D. W. and R.P. Kishan, (1998). An evaluation of Federal Reserve forecasting: Reply, *Journal of Macroeconomics*, 21(1), 189-203.

Joutz, F. and H. Stekler, (2000). An evaluation of the predictions of the Federal Reserve, *International Journal of Forecasting*, 16, 17-38.

Lahiri, K. and X. Sheng, (2008). Evolution of forecast disagreement in a Bayesian learning model, *Journal of Econometrics*, 144, 325-340.

Lahiri, K. and X. Sheng, (2010). Measuring forecast uncertainty by disagreement: The missing link, *Journal of Applied Econometrics*, 25, 514-538.

Loungani, P., H. Stekler, and N. Tamirisa, (2013). Information rigidity in growth forecasts: Some cross-country evidence, *International Journal of Forecasting*, 29(4), 605-621.

Mincer, J. and V. Zarnowitz, (1969). The evaluation of economic forecasts, In J. Mincer (Eds.) Economic Forecasts and Expectations, (National Bureau of Economic Research, New York).

Nordhaus, W. D., (1987). Forecasting efficiency, concepts, and applications, *Review of Economics and Statistics,* 69, 667-674.

Patton, A. J. and A. Timmermann, (2012). Forecast rationality tests base on multi-horizon bounds, *Journal of Business and Economic Statistics*, 30(1), 1-17.

Romer, C. D. and D. Romer (2000). Federal Reserve information and the behavior of Interest Rates, *American Economic Review*, 90(3), 429-457.

Sims, C. A., (2002). The role of models and probabilities in the monetary policy process, *Brookings Papers on Economic Activity*, 33(2), 1-62.

Sinclair, T. M., F. Joutz, and H. O. Stekler, (2010). Can the Fed predict the state of the economy? *Economics Letters,* 108(1), July, 28 – 32.