Where is the Fed in the distribution of Forecasters? November 6, 2012

By

Edward N. Gamber
Department of Economics
Lafayette College
Easton Pa 18042
610-330-5310
gambere@lafayette.edu

Julie K. Smith
Department of Economics
Lafayette College
Easton Pa 18042
610-330-5301
smithjk@lafayette.edu

and
Dylan C. McNamara (Ph.D. Candidate)
Department of Economics
The University of Michigan
Ann Arbor, MI 48109
mcnamard@umich.edu

Fields: monetary policy, forecasting

Abstract
Previous research comparing the Fed's Greenbook forecasts with a median forecast from a private-sector panel has found that the Fed's forecasts are superior. These comparisons potentially miss information from other parts of the distribution of forecast errors. We compare the Fed's forecast errors to the upper and lower quartiles from the Survey of Professional Forecasters' forecast errors and find that errors in the lower quartile are significantly smaller. We further investigate whether the forecasters who produced those forecast errors can be identified ex-ante and find that while possible the practicality of this finding is limited due to forecaster turnover.

Where is the Fed in the Distribution of Forecasters?

1. Introduction

Previous researchers have compared the accuracy of the Federal Reserve Greenbook forecasts with that of private-sector forecasters.[1] In general, those studies find that the Fed's Greenbook forecasts for real growth and inflation are significantly better (have smaller root mean squared errors) than private sector forecasts.[2,3] Most often, private-sector forecasts are represented by the mean or median of a group of forecasters such as the Survey of Professional Forecasters (SPF) or Blue Chip (BC). In this paper, we look at other parts of the distribution of private sector forecasts in order to assess whether there is some part of that distribution that is consistently as good as, or better than, the Fed. This question is of policy interest because Greenbook forecasts, which are a key part of the monetary policy decision process, are released with a five year lag. It may not be feasible for the public to construct a large econometric model of the economy to forecast growth and inflation. In addition, the public may not be able to replicate the judgmental part of the forecast produced by the Fed or a private forecaster, which they gained via repetition. If the public can use others' forecasts, which are similar to the Greenbook, to deduce the Greenbook forecast it may lead to a better understanding of monetary policy decisions.

For each period in our sample we divide the SPF forecast errors into quartiles and measure the Fed's forecast accuracy relative to the upper and lower quartiles in the distribution. We find that the Fed consistently beats the forecasters who produce errors in the upper quartile.

---

[1] See Romer and Romer (2000), Joutz and Stekler (2000), D'Agostino, et al. (2006), Faust and Wright (2007), Reifschneider, D. and P. Tulip (2007), Sinclair, Stekler, and L Kitzinger (2010), Gamber and Smith (2009).

[2] In contrast to the findings of most researchers with respect to inflation and real output growth, Baghestani (2008) finds that the private sector forecast of the unemployment rate are superior to the Federal Reserve's forecast of the unemployment rate.

[3] In Gamber and Smith (2009) we found that the RMSE's of inflation forecasts of the Fed and the private sector have moved closer to each other since the mid-1990s but the Fed RMSE's are still significantly smaller.

But forecasters who produce errors in the lower quartile are as good as, or better than, the Fed over the entire sample (1968:4 through 2002:4) as well as various sample splits and various horizons.

We also examine the behavior of joint forecast errors using two different (but standard) weighting schemes. We find that those forecasters whose joint errors are in the lower quartile are consistently as good as, or better than, the Fed over the entire sample, various sample splits and various horizons and that errors in the upper quartile are consistently worse than the Fed's.

We explore the possibility of identifying ex-ante a consistent group of forecasters, based on their past forecasting performance, who are as good as or better than the Fed. In particular, we consider the two year track record of forecasters to determine whether there is a pattern of attrition based upon the inaccuracy of their forecasts. We find that the percentage of forecasters that outperforms our benchmark ARMA model is relatively constant over time for output, but increases over time for inflation forecasts. We find little evidence of survivor bias; that is, higher forecast errors do not correlate with an increased probability to 'drop-out' of the sample.

Finally, we determine the feasibility of developing a modified consensus forecast for each of the variables, and for each of the forecast horizons. We find a small sample of forecaster ID numbers associated with superior forecasts. We successfully create a modified consensus that over a limited out of sample comparison is statistically as good as or better than the Fed. However, given the large amount of turnover in the sample, the practical applicability of this modified consensus is limited.

This paper is organized as follows. Section 2 briefly reviews related literature. Section 3 discusses the difference between ex-post and ex-ante accuracy. Section 4 describes the data. Section 5 presents the empirical results of the RMSE comparisons. Section 6 explores the

changing composition of the SPF and the possibility of survivor bias. Section 7 discusses the task of identifying individual successful forecasters and developing a modified consensus and Section 8 concludes.

2. Related Literature

Romer and Romer (2000) compared the Fed's Greenbook forecasts with several private sector forecasts. They find that for both inflation and output growth, the Fed's forecasts are superior to private sector forecasts. By superior they mean that given a Fed forecast, there is no additional information provided by private sector forecasts. Their main goal was to explain why a rise in the federal funds target is typically accompanied by a rise in long-term interest rates. Their explanation is that when the Fed raises the federal funds rate target, it implicitly reveals information about its forecast for inflation (namely, it is higher than previously thought) and this additional information leads to higher long-term rates through the standard Fisher effect.

Several papers have re-examined Romer and Romer's results. Faust, Swanson and Wright (2004) directly investigated whether the private sector revises its forecasts in response to Fed policy surprises. With the exception of Industrial Production, their answer is no, private sector forecasters do not appear to systematically incorporate monetary policy surprises into their forecasts. They take this as evidence that Fed forecasts do not contain superior information. Their methodology differs from the Romer and Romer methodology in that they use monetary policy surprises rather than the change in monetary policy. Faust, Swanson and Wright (2004) suggest that Fed forecasts may have contained superior information in the early 1980s but that informational advantage appears to have disappeared.

A related strand of literature looks at changes in the overall forecastability of the economy. Atkeson and Ohanian (2001) show that the Philips curve does a poorer job of

forecasting inflation than the naïve (random walk) model after the mid-1980s. Stock and Watson (2007) investigate why this happened and find that the reduction in volatility of inflation after the mid-1980s (the onset of the Great Moderation) is best characterized as a reduction in the volatility of the predictable component of inflation. Consequently, forecasters have lost their relative advantage over the naïve model.

D'Agostino et al. (2006) extend Stock and Watson's study to include several additional macroeconomic time series. They find that both the Fed's and the SPF's forecasts lost ground relative to the naïve benchmark forecast after the mid-1980s. They show that models that rely on cross correlations performed worse after the mid-1980s because those correlations diminished significantly with the Great Moderation.

Gamber and Smith (2009) follow up on Stock and Watson (2007) by looking at the relative forecasting accuracy of the Fed and the private sector (represented by the Blue Chip and the SPF). Stock and Watson's results suggest that it has become harder for forecasters to provide value-added over the naïve model. Thus, the superior forecasting abilities of the Fed that were identified by Romer and Romer (2000) would be expected to diminish as well. Gamber and Smith find that the Fed's forecasting advantage over the private sector did diminish after the Great Moderation and again after the Fed took steps to increase its transparency in the mid-1990s. Although the gap between the accuracy of the Fed and the private sector has narrowed, the Fed's errors still remain significantly smaller.

Crowe (2010) has shown that consensus forecasts are inefficient, even when allowing for the individual forecasts to be efficient. This results from over-weighting the prior when the average or median of a group of forecasts is calculated. Crowe's results suggest that, while each individual forecaster is communicating their best guess, they are not communicating their

idiosyncratic information, which one would need in order to create a good consensus forecast. Crowe explores alternatives to improving consensus forecasts using linear combinations of the consensus and the prior. Most importantly, Crowe's work indicates that there is significant information contained in the individual forecasts that is lost when aggregated into a consensus.

Engelberg et al. (2010) analyze the changing composition of the SPF in order to warn against traditional aggregate time-series analysis, instead suggesting focusing on a sub-panel of fixed composition, which eliminates the changes in the distribution encountered when analyzing the entire panel, for which forecasters have only an 83% one year return rate over their sample (1992:1-2006:4). While their analysis focuses on establishing the changing nature of the SPF panel, we create a modified consensus from a small, fixed sub-panel. Our approach also differs from Capistran and Timmerman (2009) who present methods for aggregating unbalanced panels while attempting to create an aggregate forecast with high information content. They note that because of the unbalanced panel structure of survey data, the inability to estimate full real-time covariance between experts' forecasts implies that other aggregations are better. In particular, they find that projecting the outcome variable on a constant and the equal-weighted forecast, which uses the full set of individual forecasts and adjusts for bias and noise, performs particularly well over their sample (1979:1 – 2006:4).

3. Ex-post versus Ex-Ante Accuracy

The above review of existing literature suggests that there has been an ongoing interest in measuring the Fed's forecast accuracy, the private-sector's forecast accuracy and the accuracy of both relative to each other and relative to a naïve benchmark. Part of the interest in these comparisons stems from the fact that the Fed's monetary policy decisions rely on forecasts and so a general assessment of the accuracy of those forecasts provide important feedback to staff

forecasters. But another closely related issue is that the Fed's Greenbook forecasts are not available to the public, except with a 5-year lag. To the extent that the Fed is more accurate, their forecast could potentially provide valuable information to private forecasters as well as consumers of forecasts, both public and private. The Fed has moved toward greater transparency since 1994, and in late 2007 they began releasing FOMC forecasts on a quarterly basis. But the staff, or Greenbook forecasts, are still kept secret in real-time.

Thus, our first goal is to determine whether there is a group of forecasters who are consistently as good as or better than the Fed. And if so, our second goal is to see whether it possible to use that group's forecasts to proxy the Fed's Greenbook forecasts in real time.

To address our first goal we compare the relative forecast abilities of the Fed and the representative best and worst forecasters from the SPF in each time period. We draw the best forecaster from the lower quartile of forecast errors and the worst forecaster from the upper quartile of forecast errors (the details are explained below). The best and worst forecasters are identified ex-post, that is, after the actual (real-time) value of the variable being forecasted has been observed.

To address our second goal we must determine whether the best forecasters in the SPF are ex-ante identifiable, based on their recent forecast performance. As a part of this analysis, it is important to understand the nebulous composition of the forecasting group. A potential pitfall of such analysis is that there could be some form of survivor bias introduced into the distribution over time. By survivor bias we mean that forecasters, who are evaluated on forecast performance, would be expected to survive as part of the distribution for a longer time if the forecasts produced are generally more accurate.

It is important to distinguish between the distribution of forecasts, and the distribution of forecasters. In the empirical work to follow, we make use of both concepts. In section 5 we focus on the distribution of forecasts and compare the Fed's forecast errors to the quartiles of errors produced by the forecasters in the SPF, not accounting for the changing composition of the sample. This analysis of the distribution of forecasts is similar to the work of previous researchers in that we ignore the changing composition. But it differs in that we compare the Fed's performance to the quartiles of the private sector distribution rather than just the mean or median. In sections 7 and 8 we focus our attention on the distribution of forecasters by looking for factors that predict whether a forecaster will be in the lower quartile of forecast errors.

4. Data

In the empirical work that follows we compare the forecast errors from the Fed's Greenbook to the forecast errors generated by two measures of the distribution of forecasts from the SPF for both inflation and real output growth. We examine the forecast errors at the boundary of the upper quartile and the lower quartile. Specifically, at each forecast date we arranged the absolute value of the forecast errors, divided the arranged sample by 4 and took the forecast at the boundaries of the third and fourth quartile (known as high quart) and the first and second quartile (known as low quart).

The Greenbook forecasts are available monthly from 1968:11 through 1980:12 and eight times a year from 1981:01 through 2002:12 for a total of 325 observations.[4] The SPF data are quarterly beginning in 1968:04 and continuing through the end of the Greenbook sample 2002:04 for a total of 137 observations. In all of our forecast error comparisons we are interested

[4] The full data set extends through 2004. We used the 2003-2004 observations to conduct out-of-sample tests (see section 7).

in comparing the errors generated by different forecasters for identical quarters.[5] To do this, we create a monthly matched data set to compare as best as possible forecasts that were based on the same information sets. Therefore, in comparing the SPF with the Fed we consider only the forecasted months for which both sets of forecasts exist.

All forecast errors are defined as "actual" minus the forecasted value where actual is either the third, or final, release of the relevant measure.[6] We consider forecast horizons 0 through 4 quarters.

5. RMSE Comparison

Our goal is to determine whether it is possible to construct a proxy for the Fed's Greenbook forecast. Because the Fed relies heavily on Greenbook forecasts when making monetary policy decisions, finding such a proxy would give us insight into the monetary policy process. The first challenge in constructing such a proxy is to determine whether there exists a group of private sector forecasters that are more accurate than the Fed. In this section, we conduct several forecast comparisons in order to determine whether such a group exists. But we do this first without considering the changing composition of the forecasters in the group. The changing composition of the forecasters is addressed in section 7.

Tables 1 and 2 show the root mean squared errors (RMSEs) for the Fed, high quart SPF and low quart SPF, over various sub-samples for inflation and real GDP growth for the monthly matched data. The RMSE is a standard measure of forecast accuracy. The full sample is 1968-

[5] We arrange our forecast data as described in Romer and Romer (2000), pp. 431-33. The Greenbook forecasts are aligned with the month that each is published. The SPF forecasts are aligned with the middle month of each quarter.
[6] We use the GNP price deflator prior to 1992. Between 1992 and 1996 we use the GDP implicit price deflator and after 1996 we use the GDP price index. We use real GNP prior to 1992. Between 1992 and 1996 we use real GDP and after 1996 we use chain-weighted real GDP. All real time data were obtained from the St. Louis Federal Reserve Web-site (http://alfred.stlouisfed.org/).

2002 for the SPF.[7] We calculated the RMSE for sample splits at 1974:09 (the date at which the BEA switched from releasing two GDP estimates (15- and 45-day) to three GDP estimates (15-, 45- and 75-day) after the end of each quarter and 1988:06 (the date at which the BEA switched from releasing the GDP estimates at 15-, 45-, and 75-day intervals to later in the month).[8] We employed the modified Diebold-Mariano test statistic to test whether the forecast errors (RMSE) were different across forecasters.[9]

In Table 1 the results for inflation are presented. The Fed is a superior forecaster to the representative worst forecaster (high quart) at all horizons and the Fed is an equivalent or slightly worse forecaster than the representative best forecaster (low quart). For real output growth, the results in Table 2 suggest that the representative best forecaster (low quart) is superior to the Fed at most time horizons. The exception is the 1968-1974 time period when the samples are extremely small. The representative worst forecaster (high quart) is worse than the Fed at all time horizons. Overall the inflation and growth results indicate that there are some forecasters who are at least as good as the Fed ex-post.

Recent research has highlighted the importance of jointly evaluating a group of forecasts[10]. Given the Federal Reserve's dual mandate, it is likely that their loss function includes both inflation and output growth. We examine whether a joint evaluation of growth and

[7] All samples end at 2002 because the Greenbook forecasts are available with a 5-year lag.
[8] See Young(1992) for more details on how the BEA's releases have changed over time.

[9] According to Harvey, Leybourne and Newbold (1997), the unmodified Diebold-Mariano test statistic is "quite seriously oversized for moderate number of observations." They suggest the following modification which results in an improvement in the behavior of the test statistic for moderately-sized samples:

$S_1^* = S_1\sqrt{\frac{T+1-2(h+1)+\frac{h(h+1)}{T}}{T}}$, $S_1 = \frac{\bar{d}}{\sqrt{\hat{V}(\bar{d})}}$, where $\bar{d}$ is the mean difference of the prediction errors and $\hat{V}(\bar{d})$ is the estimated variance. The modified Diebold-Mariano test statistic is estimated with Newey-West corrected standard errors that allow for heteroskedastic autocorrelated errors.

[10] See Gamber and Hakes (2005) and Sinclair et. al (2010).

inflation forecast errors affects the general conclusions of the previous section. We examine joint forecast errors using two variations of a standard weighting scheme. Our joint forecast errors were calculated as follows: $e_t^{p,y} = \omega_p e_t^p + \omega_y e_t^y$ where e_t^p and e_t^y are the inflation and output growth forecast errors, and $e_t^{p,y}$ is the joint forecast error. The weights ω_p and ω_y are calculated as follows:

$$\omega_p = \frac{\sigma_p^2 - \sigma_{p,y}}{\sigma_p^2 + \sigma_y^2 - 2\sigma_{p,y}} \qquad \omega_y = \frac{\sigma_y^2 - \sigma_{p,y}}{\sigma_p^2 + \sigma_y^2 - 2\sigma_{p,y}}$$

where σ_p^2 is the estimated variance of the inflation forecasts, σ_y^2 is the estimated variance of growth forecasts and $\sigma_{p,y}^2$ is the estimated covariance of inflation and growth forecasts. We calculated joint errors using two methods and found similar results. Method 1 calculates the forecast errors as the absolute forecast errors differenced from the median absolute error. It includes the Fed as a forecaster, hence subjecting Fed forecasts to the same weighting as SPF forecasts. Differencing the absolute errors from the median controls for changes in the distribution over time.[11] The weights for method 1 are shown in Table 3.

For method 2 we defined the forecast error as the raw absolute errors. We also separated the Fed from the SPF and gave each its own weighting. Across the two methods the weights are roughly the same. Both weighting schemes (Tables 4 and 5) weight output more, so we expect (and find) that the location of the Fed in the distribution of forecasters based on joint forecast errors is similar to the result we found for output growth forecast errors (Table 2).

In particular, Tables 6 and 7 show that for joint forecast errors, the Fed is consistently better than the forecasts in the higher quartile. The Fed is also generally no better than or even

[11] See Gamber, Smith and Weiss (2011) for more details on shifts in the distribution of forecast errors.

somewhat worse than the forecasts in the lower quartile, except for the subsample from 1968:11-1974:09 where there are very few observations.

6. Survivor Bias

In this section we examine whether the movement of forecasters into and out of the SPF shifts the distribution of forecasts, possibly resulting in bias for the forecasters who remain (survive) in the sample. Theoretically, the exit of certain forecasters could result in a bias in either direction. If inaccurate forecasters face a higher probability of leaving the SPF compared to accurate forecasters, we would expect such survivor bias to lead to greater accuracy of private sector forecasts time. Alternatively, if private sector forecasters are motivated more by reputational factors as in Lamont (2002) we would expect to see less accuracy over time.[12] Regardless of the direction, however, the existence of such bias would affect our ability to identify a proxy for the Fed's Greenbook forecasts.

We conducted two tests for survivor bias. For both tests we constructed a measure of pre-exit relative forecast errors for each participant that experienced an exit spell of 8 consecutive quarters or longer. The pre-exit relative forecast error is defined as the difference between the absolute value of the individual forecaster's error and the absolute value of the median forecast error averaged over the 8 quarters prior to that forecaster exiting the survey. In the first test, we regressed the pre-exit relative error on a constant. Under the null hypothesis that the exiting forecasters have errors that are no different from those remaining in the survey, the constant is zero. We find that for both inflation and output growth at all horizons, we reject the null; the pre-exit (absolute difference from median) errors are significantly greater than zero, indicating that the average pre-exit error is greater than the median error. Upon closer

[12] In fact, based on Lamont's (2002) results, if the proportion of "older and more established" forecasters in the SPF grows over time, we would expect the accuracy of the SPF decline.

examination, however, it appears that the OLS results were dominated by a few outliers. Once we control for the outliers we are no longer able to reject the null of no difference in relative errors prior to exit.

Our confidence in the outlier-adjusted result was bolstered by our second test. In our second test for survivor bias we estimated a logit regression of the form: $d_i = c + \beta x_i + \eta_i$ where d is a dummy variable that takes a value of *1* if the forecaster exits the next period and *0* otherwise, c is a constant and x_i is the forecaster i's relative forecast error. The logit results are not consistent with survivor bias; for both inflation and output growth, and for all forecast horizons, $\hat{\beta}$ is not significantly different from zero. These results were unaffected by the elimination of outliers.[13] In sum, previous forecast performance does not appear to influence whether a forecaster exits the Survey of Professional Forecasters.

7. Identification of Superior Forecasters

For a forecast proxy to be useful, one must be able to separate accurate forecasters from the inaccurate forecasters ex ante, before the data that are being forecasted are published. Romer and Romer (2000) found the Fed to be more accurate than the private sector. Using a more recent sample Gamber and Smith (2009) found the Fed is still more accurate than the private sector after 1994, but since that time the difference between the Fed's forecast accuracy and the private sector's forecast accuracy has narrowed. Using the median SPF forecast for inflation or output growth is therefore more accurate now compared to before 1994, but still less accurate than the Fed. The next logical step in this research program is to determine whether the forecasts in the SPF can be partitioned or combined in a way that improves accuracy even further, perhaps

[13] Specific model scenarios available upon request. The models do not indicate probabilities of exit greater than about 20% anywhere in the range of historical data for average forecast errors. Further, the vast majority of the distribution consists of predicted values within the range of 5% to 10%, with no apparent differentiation corresponding with exit.

enough so that the forecast errors of this select group are indistinguishable from the Fed's forecast errors.

We seek to identify a sample of forecasters for both inflation and output growth at each horizon with higher than median accuracy over the sample. We then use this subset of forecasters to create a "consensus" forecast that potentially provides a better outlook than the median of the full sample. Because we are using the forecaster ID numbers to identify and keep track of forecasters over time, we restricted the sample to 1990:3 to 2002:4, which coincides with the Philadelphia Fed managing the survey.

Utilizing the quartiles created earlier, we classify an individual forecast as "good" if it appears in the top quartile, i.e. the forecast error is in the smallest quartile. We then calculate the number of "good" forecasts that each forecaster produces, as a percent of their total number of forecasts. We use a Poisson regression[14] to determine which forecasters have more "good" forecasts than is expected. The forecasters whose number of "good" forecasts is above the two-sided 95% confidence interval for the expected value of "good" forecasts are classified as above average for that variable and horizon. Statistically close to 60-70% of forecasters can be classified as above average according to the regression. In addition, the forecasters who are "good" for each variable and horizon vary greatly, with very little overlap across variables and only slightly more overlap across horizons within variables.

To construct a "consensus" forecast from our sample of good forecasters, we restricted our sample as follows. We considered only those forecasters who participated in the survey

[14] Technically, we have a binomial distribution. However, as we seek to predict the expected number of good forecasts, rather than the percentage, it is appropriate to use the limiting Poisson distribution (Greene (2008)).

more the median number of times.[15] From that sub-sample, we chose only those forecasters who appeared in the top quartile at least 40% of the time. The modified consensus is the median of these forecasters' forecasts (referred to as the Median of the Modified Consensus, or MMC).

Tables 8 and 9 display results from our out-of-sample forecasting exercise using the modified consensus. The out-of-sample period was 2003:1 to 2004:4. In general, the modified consensus is not statistically different than the median SPF forecast, except for current period inflation. Also, for inflation forecasts, the Fed performed worse than the SPF median and worse than the modified consensus over this sample. For output, all three forecasts were statistically indistinguishable. As a result, it seems that this is a methodology that, at least for the given sample, identifies a subset of forecast IDs that can be used to approximate forecasts as good as or better than the Fed's.

The feasibility of such a consensus is brought into question as many forecasters drawn from the sample 1990:3 to 2002:4 as successful had their success earlier in the sample and have stopped responding (or forecasting) by the start of the out of sample comparison, severely restricting the variation for the new consensus. In order to account for this issue, it may be possible to create a rolling sample for the "good" forecasters based on a shorter history and updating each quarter. Furthermore, with only a limited out of sample comparison, it is possible the results are sensitive to the time period, during which the Fed was notably concerned with low inflation. Also, since there are only three monthly-matched data points, we have used quarterly-matched. This places the Fed at an information disadvantage in quarters where their forecast is in the first month of the quarter.

[15] This value varies slightly by variable and horizon, but is between 10 and 12 forecasts (out of a possible 54). We make an exception for those forecasters whose fewer than the median forecasts are concentrated in the period 2000:1 to 2002:4 as this allows forecasters that are new in this period to be included.

Our difficulty in attempting to formulate a modified consensus using this approach is consistent with analysis of the composition of the SPF by Engelberg et al. (2010). Their analysis is undertaken on data from 1992 to 2006. Over this period they find an average of 33.7 forecasters per quarter. They find an average of 9.8 forecasters participating in the current survey who did not participate four quarters earlier and 8.9 forecasters not participating in the current survey who did participate four quarters earlier. Due to the rapid changes in the composition of the panel, they recommend restricting any aggregate analysis to sub-panels with fixed composition.

8. Conclusion, Policy Implications and Issues for Further Study.

This paper addresses two questions: 1) is there a group of forecasters that is as good or better than the Fed at forecasting inflation and real output growth and 2) is there a way to identify those forecasters? In response to the first question we see that the representative best forecaster is as good as the Fed for both inflation and real output growth, but that the Fed is consistently better than the representative worst forecaster. This result is robust to considering joint forecasts in which inflation and output forecasts are considered a single jointly made forecast; there is still a representative best forecaster that is as good as the Fed.

In considering methods to identify these better forecasters, we first test for potential biases in the distribution due to entry and exit from the survey. We find no evidence of survivor bias; forecasters with lower errors are not more likely to remain in the survey. We are able to identify a group of forecasters that perform as good as or better than the Fed for each variable and at each forecast horizon. These forecasters continue to perform well out of sample. The practical applicability of this improved consensus is limited by the large amount of turnover in the Survey of Professional Forecasters.

When making consumption and investment decisions, the private sector relies on signals from the Federal Reserve about the future course of monetary policy. Since 1994, the Federal Reserve has moved toward greater transparency.[16] Despite the increased transparency however, the Fed continues to release its Greenbook forecasts with a five-year lag. Given the importance of the Greenbook forecasts in the Fed's policy decision process, the lag in releasing these forecasts contributes to a continued source of policy opacity for the public.

Our research suggests the lag in releasing the Greenbook forecasts may be a non-issue for a certain sub-set of forecasters who consistently produce forecasts that are as accurate, and in some cases more accurate, than the Fed's forecasts. The next step in this research program is to further refine the criteria for selecting, ex ante, the forecasters that might populate the sub sample from which an improved consensus forecast is derived.

[16] Other studies such as Croushore and Koot (1994) and Casillas-Overa et. al (2006) have examined transparency and credibility of the central bank by comparing private sector and central bank forecasts.

References

Atkeson, A. and L. E. Ohanian, 2001. "Are Phillips Curves Useful for Forecasting Inflation?" *Federal Reserve Bank of Minneapolis Quarterly Review*, vol. 25 (2001), 2-11.

Baghestani, Hamid. 2008. "Federal Reserve versus private information: Who is the best unemployment rate predictor?" *Journal of Policy Modeling*: vol. 30.1 pp. 101-110.

D'Agostino, A., D. Giannone and P. Surico, 2006. "(Un)predictability and Macroeconomic Stability," European Central Bank working paper, no. 605 (2006).

Capistran, Carlos and Allan Timmerman, 2009. "Forecast Combination with Entry and Exit of Experts," *Journal of Business and Economic Statistics*, 27, 429-440.

Casillas-Olvera, G. and Bessler, D. A. (2006) Probability forecasting and central bank accountability, *Journal of Policy Modeling*, vol. 28, pp. 223-34.

Croushore, D., & Koot, R. S. (1994). A measure of Federal Reserve credibility. *Journal of Policy Modeling*, vol. 16, pp. 215–31.

Crowe, Christopher, 2010, "Consensus Forecasts and Inefficient Information Aggregation, International Monetary Fund working paper 10/178.

Engelberg, Joseph, Charles F. Manski, Jared Williams, 2010. "Assessing the Temporal Variation of Macroeconomic Forecasts by a Panel of Changing Composition," *Journal of Applied Econometrics*, n/a. doi: 10.1002/jae.1206

Faust, J. and E. T. Swanson and J. H. Wright. 2004 "Do Federal Reserve Policy Surprises Reveal Superior Information about the Economy?," *Contributions to Macroeconomics*: Vol. 4 : Issue 1, Article 10.

_____, 2007. "Comparing Greenbook and Reduced Form Forecasts using a Large Realtime Dataset," NBER working paper 13397 (2007).

Gamber, E. N. and D. R. Hakes. 2005. "Is monetary policy important for forecasting real growth and inflation?" *Journal of Policy Modeling*. Vol 27.2 pp. 177-187, 2005.

_____ J. K. Smith 2009. "Are the Fed's Inflation Forecasts Still Superior to the Private Sector's?" *Journal of Macroeconomics*, vol. 31, pp. 240-51.

_____ and M. Weiss, 2011. "Forecast Errors Before and After the Great Moderation," *Journal of Economics and Business*, vol. 63, no. 4, pp. 278-89.

Greene, William H. *Econometric Analysis*. 6th ed. Upper Saddle River, N.J.: Pearson Prentice Hall, 2008.

Harvey, David, Stephen Leybourne and Paul Newbold, 1997. "Testing the equality of prediction mean squared errors." *International Journal of Forecasting,* vol. 13 (1997) 281-291.

Joutz, F. and H.O. Stekler, (2000). "An Evaluation of the Predictions of the Federal Reserve," *International Journal of Forecasting,* vol. 16, 17-38.

Lamont, O. A., 2002, "Macroeconomics Forecasts and Microeconomic Forecasters," *Journal of Economic Behavior and Organization*, v. 48, no. 3, July 2002, pp. 265-80.

Reifschneider, D. and P. Tulip, 2007. "Gauging the Uncertainty of the Economic Outlook from Historical Forecasting Errors," Federal Reserve Board of Governors Economic and Finance Discussion Series, no. 2007-60 (2007).

Romer, C. and D. Romer, 2000. "Federal Reserve Information and the Behavior of Interest Rates," *American Economic Review*, vol. 90 (2000) 429-57.

Sinclair, T., H. Stekler, and L Kitzinger, 2010. "Directional Forecasts of GDP and Inflation: A Joint Evaluation With an Application to Federal Reserve Predictions," *Applied Economics*, vol. 42, no. 18, pp. 2289-2297.

Stock, J. and M. Watson, 2007. Why Has U.S. Inflation Become Harder to Forecast," *Journal of Money, Credit and Banking*, vol. 39.

Young, Allan H. 1992, "Alternative Measures of Change in Real Output and Prices,"*Survey of Current Business* 72 (April 1992): 32–48.

Table 1:Inflation

sample	horizon	RMSE Fed	RMSE High quart	RMSE Low quart	Ratio(High/Fed)	Ratio(Low/Fed)	Modified DM (High=Fed)		Modified DM (Low=Fed)		N
1968:11-2002:12	0	1.17	1.94	1.00	1.66	0.86	12.27	***	-2.54	**	110
	1	1.53	2.24	1.44	1.47	0.95	13.39	***	-1.28		110
	2	1.81	2.54	1.78	1.40	0.98	12.16	***	-0.87		109
	3	1.85	2.56	1.77	1.38	0.96	12.51	***	-1.17		104
	4	1.89	2.69	1.90	1.42	1.00	14.72	***	0.98		95
1968:11-1974:09	0	1.50	2.58	1.45	1.72	0.96	8.82	***	0.10		23
	1	2.53	3.40	2.47	1.34	0.98	7.95	***	0.15		23
	2	3.08	4.01	3.25	1.30	1.05	6.00	***	0.88		22
	3	3.50	4.27	3.41	1.22	0.97	4.59	***	-0.86		17
	4	4.69	5.58	4.80	1.19	1.02	3.53	***	0.50		8
1974:10-1988:06	0	1.32	2.10	1.04	1.59	0.79	7.69	***	-3.04	***	46
	1	1.29	2.16	1.20	1.67	0.93	10.29	***	-0.87		46
	2	1.66	2.48	1.42	1.50	0.86	8.15	***	-1.40		46
	3	1.64	2.58	1.54	1.58	0.94	10.27	***	-0.65		46
	4	1.69	2.73	1.64	1.61	0.97	11.66	***	0.41		46
1988:07-2002:12	0	0.66	1.17	0.54	1.78	0.82	7.58	***	-1.16		41
	1	0.88	1.32	0.75	1.50	0.85	8.13	***	-2.37	**	41
	2	0.75	1.31	0.72	1.75	0.96	9.73	***	-0.98		41
	3	0.77	1.25	0.70	1.63	0.91	9.27	***	-0.69		41
	4	0.86	1.52	0.91	1.76	1.05	10.10	***	1.06		41

Notes: * = .10, ** = .05, *** = .01.

Table 2: Output growth

sample	horizon	RMSE Fed	RMSE High quart	RMSE Low quart	Ratio(High/Fed)	Ratio(Low/Fed)	Modified DM (High=Fed)		Modified DM (Low=Fed)		N
1968:11-2002:12	0	2.09	3.27	1.69	1.56	0.81	11.95	***	-3.83	***	110
	1	3.20	3.96	2.50	1.24	0.78	9.69	***	-7.06	***	110
	2	3.39	4.33	2.91	1.28	0.86	8.36	***	-5.49	***	109
	3	3.52	4.30	3.01	1.22	0.86	9.08	***	-5.30	***	104
	4	2.98	3.93	2.61	1.32	0.88	9.75	***	-3.19	***	95
1968:11-1974:09	0	1.66	2.93	1.12	1.77	0.67	7.97	***	-1.61		23
	1	3.36	4.16	2.70	1.24	0.81	4.65	***	-2.46	**	23
	2	4.50	5.93	4.39	1.32	0.98	5.17	***	-1.31		22
	3	5.07	6.46	4.91	1.27	0.97	6.16	***	-1.69		17
	4	4.94	6.39	4.89	1.29	0.99	3.28	**	-0.13		8
1974:10-1988:06	0	2.67	4.19	2.19	1.57	0.82	8.50	***	-2.52	**	46
	1	3.89	4.86	3.03	1.25	0.78	7.27	***	-5.37	***	46
	2	3.54	4.64	2.88	1.31	0.81	6.74	***	-3.56	***	46
	3	3.83	4.61	3.05	1.20	0.80	6.70	***	-3.41	***	46
	4	3.10	4.31	2.64	1.39	0.85	9.16	***	-1.97	*	46
1988:07-2002:12	0	1.47	2.02	1.26	1.37	0.86	8.35	***	-2.51	**	41
	1	2.03	2.42	1.55	1.19	0.77	5.45	***	-4.94	***	41
	2	2.36	2.61	1.68	1.10	0.71	3.68	***	-4.86	***	41
	3	2.09	2.42	1.60	1.16	0.77	4.53	***	-4.21	***	41
	4	2.24	2.65	1.84	1.18	0.82	4.89	***	-3.27	***	41

Notes: * = .10, ** = .05, *** = .01.

Table 3: Method 1 Weights: SPF and Fed

horizon		
0	0.30	0.70
1	0.30	0.70
2	0.27	0.73
3	0.23	0.77
4	0.25	0.75

Table 4: Method 2: SPF Weights

horizon		
0	0.25	0.75
1	0.19	0.81
2	0.18	0.82
3	0.12	0.88
4	0.16	0.84

Table 5: Method 2: Fed Weights

horizon		
0	0.17	0.83
1	0.13	0.87
2	0.14	0.86
3	0.09	0.91
4	0.15	0.85

Table 6: Joint Forecasts Method 1

sample	horizon	RMSE Fed	RMSE High quart	RMSE Low quart	Ratio(High/Fed)	Ratio(Low/Fed)	Modified DM (High=Fed)		Modified DM (Low=Fed)		N
1968:11-2002:12	0	1.72	2.69	1.42	1.56	0.83	10.08	***	-3.48	***	110
	1	2.56	3.23	2.09	1.26	0.82	10.11	***	-6.24	***	110
	2	2.84	3.66	2.50	1.29	0.88	7.68	***	-5.13	***	109
	3	3.02	3.75	2.64	1.24	0.87	8.12	***	-4.90	***	104
	4	2.58	3.44	2.36	1.33	0.91	8.80	***	-2.46	**	95
1968:11-1974:09	0	1.58	2.72	1.28	1.71	0.81	8.34	***	-0.73		23
	1	3.04	3.83	2.51	1.26	0.82	6.01	***	-2.07	**	23
	2	3.99	5.19	3.86	1.30	0.97	4.92	***	-1.13		22
	3	4.55	5.73	4.38	1.26	0.96	6.52	***	-1.79	*	17
	4	4.74	5.95	4.73	1.25	1.00	2.96	**	-0.11		8
1974:10-1988:06	0	2.15	3.34	1.76	1.55	0.82	7.81	***	-2.85	***	46
	1	2.94	3.76	2.42	1.28	0.82	9.84	***	-4.22	***	46
	2	2.89	3.83	2.44	1.33	0.85	6.75	***	-3.23	***	46
	3	3.21	3.99	2.66	1.24	0.83	6.47	***	-2.84	***	46
	4	2.62	3.73	2.33	1.43	0.89	10.03	***	-1.48		46
1988:07-2002:12	0	1.15	1.66	1.00	1.44	0.87	8.30	***	-2.58	**	41
	1	1.62	1.96	1.27	1.21	0.79	5.68	***	-4.75	***	41
	2	1.89	2.14	1.39	1.14	0.74	4.20	***	-4.11	***	41
	3	1.74	2.03	1.33	1.17	0.77	5.01	***	-4.96	***	41
	4	1.82	2.21	1.54	1.21	0.85	4.80	***	-2.45	**	41

Notes: * = .10, ** = .05, *** = .01.

Table 7: Joint Forecasts Method 2

sample	horizon	RMSE Fed	RMSE High quart	RMSE Low quart	Ratio(High/Fed)	Ratio(Low/Fed)	Modified DM (High=Fed)		Modified DM (Low=Fed)		N
1968:11-2002:12	0	1.88	2.77	1.44	1.47	0.77	9.27	***	-3.96	***	110
	1	2.91	3.47	2.20	1.19	0.76	8.70	***	-6.53	***	110
	2	3.10	3.87	2.62	1.25	0.85	6.87	***	-5.62	***	109
	3	3.31	3.99	2.77	1.21	0.84	7.51	***	-5.65	***	104
	4	2.95	3.57	2.39	1.21	0.81	5.19	***	-2.69	**	95
1968:11-1974:09	0	1.62	2.76	1.25	1.71	0.77	8.17	***	-0.92		23
	1	3.22	3.94	2.54	1.22	0.79	5.02	***	-2.25	**	23
	2	4.23	5.43	4.02	1.28	0.95	4.67	***	-1.26		22
	3	4.85	6.07	4.58	1.25	0.95	6.19	***	-2.05	*	17
	4	5.93	5.74	4.39	0.97	0.74	0.05		-1.03		8
1974:10-1988:06	0	2.37	3.45	1.80	1.45	0.76	7.27	***	-3.12	***	46
	1	3.48	4.13	2.61	1.19	0.75	7.50	***	-4.59	***	46
	2	3.20	4.10	2.58	1.28	0.81	6.25	***	-3.57	***	46
	3	3.57	4.26	2.79	1.19	0.78	5.93	***	-3.41	***	46
	4	2.80	3.91	2.41	1.40	0.86	9.07	***	-1.94		46
1988:07-2002:12	0	1.29	1.71	1.04	1.33	0.81	6.70	***	-3.28	**	41
	1	1.85	2.12	1.35	1.15	0.73	4.29	***	-5.15	***	41
	2	2.12	2.29	1.46	1.08	0.69	3.10	***	-4.66	***	41
	3	1.94	2.18	1.43	1.13	0.74	4.12	***	-5.40	***	41
	4	1.98	2.33	1.63	1.18	0.82	4.16	***	-3.04	**	41

Notes: * = .10, ** = .05, *** = .01.

Table 8: Inflation, Out of Sample Comparison

sample	horizon	RMSE Fed	RMSE MMC	RMSE Median	Ratio (MMC/Fed)	Ratio (Median/Fed)	Ratio (MMC/Median)	Modified DM (MMC=Fed)		Modified DM (Median=Fed)		Modified DM (MMC=Median)		N
	0	0.89	0.89	0.71	1.00	0.80	1.25	0.81		-2.05	*	6.66	***	8
2003:1	1	1.27	0.87	0.81	0.69	0.64	1.08	-1.57		-2.26	*	0.90		8
-	2	1.27	0.85	0.71	0.67	0.56	1.19	-5.36	***	-4.16	***	0.18		8
2004:4	3	1.39	0.98	0.86	0.71	0.62	1.14	-3.18	***	-3.22	***	-0.03		8
	4	1.48	0.97	1.01	0.65	0.68	0.96	-3.84	***	-8.60	***	-0.41		8

Notes: * = .10, ** = .05, *** = .01.

Table 9: Output, Out of Sample Comparison

sample	horizon	RMSE Fed	RMSE MMC	RMSE Median	Ratio (MMC/Fed)	Ratio (Median/Fed)	Ratio (MMC/Median)	Modified DM (MMC=Fed)	Modified DM (Median=Fed)	Modified DM (MMC=Median)		N
	0	1.77	1.82	1.73	1.03	0.98	1.05	-0.49	-1.21	0.13		8
2003:1	1	1.94	1.91	1.67	0.98	0.86	1.14	-0.36	-1.68	1.51		8
-	2	1.71	1.78	1.62	1.04	0.95	1.10	-0.63	-1.07	1.25		8
2004:4	3	1.02	0.66	0.43	0.65	0.42	1.52	-1.00	-1.29	0.66		8
	4	1.22	0.91	0.73	0.74	0.60	1.25	-0.91	-1.71	2.24	*	8

Notes: * = .10, ** = .05, *** = .01.