Gauging the Uncertainty of the Economic Outlook From Historical Forecasting Errors

David Reifschneider and Peter Tulip*

August 11, 2008

Abstract

The Federal Open Market Committee (FOMC) of the Federal Reserve has recently begun publishing more information about the economic outlook, including assessments of the uncertainty attending its forecasts of real activity and inflation. Although these uncertainty assessments are qualitative in nature, they include explicit comparisons to quantitative estimates of the typical range of uncertainty facing macroeconomic projections. As we document in this paper, these quantitative estimates are based on the average forecast errors made by various private and government forecasters over the past twenty years. One implication of the estimates is that, if historical performance is a reasonable guide to the accuracy of future forecasts, considerable uncertainty surrounds all macroeconomic projections, including those of FOMC participants.

* Board of Governors of the Federal Reserve System; email addresses: david.l.reifschneider@frb.gov and peter.j.tulip@frb.gov. We thank Spencer Dale, William English, Steven Kamin, Deborah Lindner, Brian Madigan, Michael McCracken, Simon Potter, John Roberts, Glenn Rudebusch, John Williams, and Jonathan Wright for helpful comments and suggestions. The views expressed herein are those of the authors and do not necessarily reflect those of the Board of Governors of the Federal Reserve System or its staff.

1. Introduction

Since the late 1970s, the Federal Open Market Committee (FOMC) of the U.S. Federal Reserve has regularly published summaries of the forecasts of real activity and inflation made by Committee participants.[1] Recently, the FOMC expanded the amount of information it provides the public on the economic outlook in several dimensions.[2] For example, the Committee now releases projections four times of year instead of twice, has substantially increased the horizon of the projections and expanded the accompanying narrative, and provides more details on the dispersion of views among participants. The FOMC also has begun providing more information on the uncertainty associated with the economic outlook. Specifically, the Committee now releases a summary of participants' qualitative assessments of how the current level of uncertainty compares with a historical benchmark of the typical level of uncertainty associated with macroeconomic projections. This summary also reports participants' views on whether the risks to the outlook for real activity and inflation are skewed in one direction or another.

The purpose of this paper is to document and discuss the historical benchmark estimates of average forecast uncertainty that Committee participants now employ in assessing current uncertainty. To this end, we provide evidence on the past predictive accuracy of a number of different forecasters, including FOMC participants, the staff of the Federal Reserve Board, the Congressional Budget Office, the Administration, the Blue Chip consensus forecasts, and the Survey of Professional Forecasters. We then discuss how these various measures of historical predictive accuracy are used to provide the specific uncertainty benchmarks employed by the FOMC. To preview our main results, we find that:

---

[1] The Federal Open Market Committee consists of the seven members of the Board of Governors of the Federal Reserve System, the president of the Federal Reserve Bank of New York, and, on a rotating basis, four of the remaining eleven presidents of the regional Reserve Banks. In this paper, the phrase "FOMC participants" encompasses the seven members of the Board and all twelve Reserve Bank presidents because all participate fully in FOMC discussions and all provide individual forecasts; the *Monetary Policy Report to the Congress* and the *Summary of Economic Projections* provide summary statistics for their nineteen projections. (From time to time, vacancies on the Board of Governors mean that the actual number of individual forecasts is somewhat smaller.)

[2] A general discussion of these changes is in Bernanke (2007). The Committee publishes this information quarterly in the *Summary of Economic Projections* that it provides with the release of the minutes of FOMC meetings. The Federal Reserve Board's biannual *Monetary Policy Report to the Congress* also provides a summary of this information.

- Historical forecast errors are large in economic terms, indicating that — if past performance is a good guide to future accuracy — uncertainty about the economic outlook is considerable.
- Average differences in predictive performance across the forecasters in our sample are small, suggesting that the estimates are not very sensitive to the composition of our panel of forecasters. This result further implies that, in situations where data are not available at certain horizons for certain series for all the forecasters in our sample, we can produce reasonable benchmark estimates of forecast uncertainty using the subset of forecasters for which such information is available.
- About 70 percent of historical outcomes have fallen within one root mean squared error of forecasts. This result implies that historical prediction errors provide a reasonable basis for making explicit probability statements about the accuracy of future projections, conditional (once again) on the past being a good guide to future conditions.

Before discussing how we arrive at these conclusions, we first should explain why such estimates of historical forecasting accuracy are a useful benchmark against which to gauge the uncertainty of the economic outlook.

All macroeconomic projections are subject to error. The economy may be hit with any number of unforeseen developments. The available measurements of real activity and inflation may be flawed. And models of the economy may be misspecified in critical ways. For all these reasons, the likelihood that actual outcomes will deviate substantially from predicted values is considerable. This likelihood does not mean that macroeconomic projections are worthless; rather, it implies that point forecasts of real activity and inflation provide an incomplete picture of the economic outlook. Achieving a more complete picture requires additional information about the probability distribution of the various possible outcomes. In particular, we would like to be able to make statements of the form: “There is a 70 percent probability that actual GDP growth next year will fall between X percent and Y percent and a 70 percent probability that consumer price inflation will fall between A percent and B percent.”

A forecaster who wishes to make such a probability statement has several options for obtaining the necessary information.[3] The option pursued in this paper is to look to past prediction errors as a rough guide to the magnitude of forecast errors that may occur in the future. For example, if 70 percent of actual outcomes over history fell within a band of a particular width around the predicted outcomes, then a forecaster might expect future outcomes to cluster around his or her current projection to a similar degree. Such an error-based approach has two attractive features. First, the relationship of the uncertainty estimates to historical experience is clear. Second, the approach focuses on the actual historical performance of forecasters under true "field conditions" and does not rely on after-the-fact analytic calculations, using various assumptions, of what their accuracy might have been.

Admittedly, the error-based approach has a potential drawback: It assumes that the past is a good guide to the future. Although this assumption in one form or another underlies all statistical analyses, there is always a risk that structural changes to the economy may have altered its inherent predictability, thereby reducing the relevance of past forecasting performance. Indeed, recent studies by Tulip (2005) and Campbell (2007) find that a statistically and economically significant reduction in the size of macroeconomic forecast errors occurred in the mid-1980s. We see this evidence as a reason for being wary about looking too far back in time for guidance, not an across-the-board invalidation of the relevance of past experience to gauging future uncertainty. That said, these studies suggest the need to be alert to evidence of structural change and other factors that may alter the predictability of economic outcomes for better or worse.

Our implementation of the error-based approach involves measuring the average accuracy of forecasts over history. A limitation of this procedure is that it provides guidance only on the average degree of uncertainty seen in the past and not on how a forecaster today, after taking account of conditions as they now stand, may perceive the degree to which uncertainty currently deviates from historical norms. To address this concern, each FOMC participant now provides a qualitative assessment of whether the uncertainty attending his or her current projection, taking account of the current situation, is greater, smaller or broadly similar to the average of the past. In making this judgment,

[3] For a general review of this topic, see Tay and Wallis (2000).

participants use the estimates of past average forecasting accuracy reported in this paper to gauge the “typical” degree of uncertainty. Participants, if they choose, also note specific factors influencing their assessments of uncertainty. For example in late 2007 and in 2008 they cited ongoing financial market strains and restricted credit availability as forces creating more uncertainty than normal about the outlook for real activity.[4] As this example illustrates, uncertainty is conditional, and perceptions of its magnitude may change from period to period in response to specific events.

Model simulations provide another way to gauge the uncertainty of the economic outlook. Given an econometric model of the economy, we can repeatedly simulate it while subjecting the model to stochastic shocks of the sort experienced in the past; for this purpose, we can use models ranging in size from simple univariate or VAR specifications to the large-scale models maintained at central banks. This approach has several advantages. For example, we can use it to approximate the entire probability distribution of possible outcomes for the economy. Moreover, we can generate these distributions as far into the future as desired and in as much detail as the structure of the model allows. Furthermore, the model-based approach permits analysis of the sources of uncertainty and can help explain why uncertainty might change over time.

However, the model-based approach also has its limitations. First, the estimates are specific to the particular model used in the analysis. If the forecaster and his or her audience are worried that the model in question is not an accurate depiction of the economy (as is always the case to some degree), they may not find its uncertainty estimates credible. Second, like the forecast-error-based approach, the model-based approach relies on the past being a good guide to the future. Finally, the model-based approach abstracts from both the difficulties and advantages of real-time forecasting: It tends to understate uncertainty by exploiting after-the-fact information to design and estimate the model, and it tends to overstate uncertainty by ignoring extra-model information available to forecasters at the time. For all these reasons, we do not emphasize the model-based approach in this paper, although we do compare our

---

[4] See the *Summary of Economic Projections* that accompanied the release of the minutes for the October 2007 FOMC meeting and for the January, April, and June FOMC meetings in 2008.

benchmark estimates of uncertainty with ones derived from stochastic simulations of FRB/US, a model used in the Federal Reserve Board for forecasting and policy analysis.

A third approach to gauging uncertainty is to have forecasters provide their own judgmental estimates of the confidence intervals associated with their projections. Such an approach does not mean that forecasters generate probability estimates with no basis in empirical fact; rather, the judgmental approach simply requires the forecaster, after reviewing the available evidence, to write down his or her best guess about the distribution of risks. Some central banks now combine judgment with other analyses to construct subjective fan charts that illustrate the uncertainty surrounding their outlooks; for example, such fan charts have been a prominent feature of the Bank of England's *Inflation Report* since the mid-1990s.

While the experience of these central banks demonstrates that subjective probability fan charts can be effective communication tools, this approach is harder to apply in the context of the FOMC. One difficulty is the lack of a single consensus forecast around which to center the distribution of possible outcomes; instead of one unified projection, the Board members and the Bank presidents produce nineteen individual forecasts. A related difficulty is the problem of summarizing how Committee participants, as a group, view uncertainty. While each Committee participant in principle may have an explicit quantitative assessment of the uncertainty surrounding his or her own outlook, the best way to aggregate such information into an informative quantitative assessment of the group's overall view is unclear and might be hard to implement. As we noted above, however, the Committee now reports participants' qualitative views on how the uncertainty of the current outlook compares with historical experience, and these assessments incorporate judgment to varying degrees.

In the remainder of this paper, we lay out a procedure for using past forecast errors to provide a benchmark estimate of historical forecast uncertainty. We begin by discussing several general considerations that influence the way in which we collect historical forecast data. We then turn to a detailed discussion of our six sources of forecast information — the FOMC, the Federal Reserve Board staff, the Congressional Budget Office, the Administration, the Blue Chip, and the Survey of Professional

Forecasters. Section 4 of the paper presents our empirical results; we then conclude with a few caveats.

2. Collecting Historical Forecast Data

To provide a benchmark against which to assess the uncertainty associated with the projections provided by individual Committee participants, one obvious place to turn is the FOMC's own forecasting record — and indeed, we exploit this information in our analysis. For several reasons, however, we also take account of the projection errors of other forecasters. First, although the Committee has provided projections of real activity and inflation for almost thirty years, the horizon of these forecasts was, for quite a while, considerably shorter than it is now — at most one and a half years ahead as compared with roughly three years under the new procedures. Accordingly, we must look to other sources to provide benchmark information on the potential accuracy of the Committee's new longer-range forecasts. Second, the definition of inflation projected by FOMC participants has changed over time in important ways, making it problematic to relate participants' past prediction errors to its current forecasts. In contrast, other forecasters have published inflation projections over many years using the same unchanged measure of consumer prices. Finally, given that the composition of the FOMC has changed over time, consideration of other forecasts reduces the likelihood of placing undue weight on a potentially unrepresentative record. For these reasons, we believe that supplementing the Committee's record with that of other forecasters has the potential to yield more-reliable estimates of forecast uncertainty.

In addition to seeking out multiple sources of forecast information, we also are interested in projections released at specific times of the year. At the time of this writing, the FOMC schedule involves publishing economic projections four times a year in conjunction with the release of the minutes of the January, April, June, and October FOMC meetings. Accordingly, we would like our forecast data to have publication dates that match this winter-spring-summer-autumn schedule as closely as possible. In principle, it would be possible to compare forecast errors from other times of the year, but we assume this would not be worth the extra effort and complication.

Note that this seasonal forecast schedule does not correspond exactly with the quarterly division of the calendar used, for example in reporting economic data. For example, some of the "summer" projections are produced late in the second quarter of the year but published early in the third quarter. It seems less confusing referring to such forecasts by season rather than quarter. However, given that the data and some forecasts are on a quarterly basis, some comparisons across the different timing schemes are unavoidable.

Under the FOMC's new communication procedures, the Committee periodically releases projections of real GDP growth, the civilian unemployment rate, total personal consumption expenditures (PCE) chain-weighted price inflation, and core PCE chain-weighted price inflation (that is, excluding food and energy). Ideally, the economic measures projected historically by our sample of forecasters would correspond exactly to these definitions; unfortunately, this has not always been the case. As discussed in the next section, the discrepancies in our sample from the Committee's new procedures are, for the most part, minor and do not have serious implications for our measures of uncertainty. However, our sample of inflation forecasts may be an exception. As noted, Committee participants now use the PCE chain-weighted price index (both overall and core) as the basis for their inflation forecasts, but over history neither the Committee nor most other forecasters consistently made inflation projections on this basis. Rather, consistent-definition inflation projections are available from a variety of forecasters over a long period for the total consumer price index (CPI). Using the accuracy of CPI inflation forecasts to gauge the uncertainty of either total or core PCE inflation raises questions of comparability because price indexes differ in volatility and hence predictability. Fortunately, the staff of the Federal Reserve Board has long produced separate inflation forecasts for all these various price measures, so we are able to compare the accuracy of CPI-based forecasts with that of projections for total and core PCE inflation.

A final issue in data collection concerns the appropriate historical period for evaluating forecasting accuracy. In deciding how far back in time to go, we face two competing effects. On the one hand, collecting more data by extending the sample further back in time should yield more accurate estimates of forecast accuracy if the forecasting environment has been stable over time. On the other hand, if the environment has in fact changed materially because of structural changes to the economy or improvements in forecasting techniques, then keeping the sample period relatively short should yield estimates that more accurately reflect current uncertainty. In balancing these two concerns, we have elected to draw forecast errors from approximately the last twenty years. Specifically, we use forecasts and outcomes from 1986 to 2006, thus providing us with 21 current-year errors, 20 next-year errors, 19 two-year-ahead errors and so forth. Tulip (2005) reports that, although forecast accuracy improved in the mid-1980s, it has not clearly changed since. Current FOMC procedures involve rolling this window forward as new data arrives. Hence, the *Summary of Economic Projections* released in May 2008 reported average forecast errors for the period 1987 to 2007.[5]

3. Data Sources

For the reasons just discussed, we have computed historical projection errors based on projections made by a variety of forecasters. Our first source is the FOMC itself, for which we employ the midpoint of the central tendency ranges reported in past releases of the *Monetary Policy Report*.[6] Our second source is the staff of the Federal Reserve Board, which prepares a forecast prior to each FOMC meeting; these projections are unofficially but universally called Greenbook forecasts.[7] Our third and fourth sources are the Congressional Budget Office (CBO) and the Administration, both of which regularly publish forecasts as part of the federal budget process. Finally, we have two private data sources — the monthly Blue Chip consensus forecasts and the mean responses to the quarterly Survey of Professional Forecasters (SPF). Both of these

[5] As discussed by McConnell and Perez-Quiros (2000) and many others, macroeconomic volatility in the United States was much larger prior to the mid-1980s; moreover, as noted above, studies have found that the size of forecast errors changed around this time. By gauging current uncertainty with data from the past twenty years alone, we are thus implicitly assuming that the calm conditions since the Great Moderation will persist into the future.

[6] Historically, the *Monetary Policy Report* has not reported the individual projections of FOMC participants but only two summary statistics — the range across all projections (generally nineteen) and a trimmed range intended to express the central tendency of the Committee's views. For each year of the projection, the central tendency is the range for each series after excluding the three highest and three lowest projections.

[7] Under FOMC confidentiality rules, individual Greenbook forecasts become publicly available only with a lag of five years. However, we are able to publish summary statistics that include the more recent forecast information, and so we use the same sample period to analyze the accuracy of Greenbook errors that we do for the other forecasters — that is, projections published from 1986 through 2006.

private surveys include a large number of business forecasters; the SPF also includes forecasters from universities and other nonprofit institutions.

Because these six sources did not project real activity and inflation in a uniform manner, they create some technical and conceptual issues for our analysis. We now discuss some of the key differences among our sources, including variations in data coverage and in reporting basis, and consider the implications of those differences. We then address several other issues important to our analysis, such as how to define “truth” in assessing forecasting performance, the mean versus modal nature of projections, and the implications of conditionality.

Data coverage

As summarized in Table 1, our data sources differ in several ways with regard to data coverage. For example, although all our sources published forecasts in every year from 1986 to 2006, only the Greenbook, the Blue Chip, and the SPF released projections near the four points of the year on the FOMC’s new publication schedule; in contrast, the FOMC, the CBO, and the Administration did not publish forecasts during the spring and autumn.[8] For this latter group, we approximate their missing projections by averaging the forecasts each made in adjacent periods; the pseudo-spring forecast is the average of their winter and summer projections, and the pseudo-autumn forecast is the average of their summer projection and following winter forecast.[9]

---

[8] Based on the 2008 and preliminary 2009 calendars, the FOMC has or will release economic projections in February, May, July and November around the middle of the month. By comparison, the forecasts in our dataset were typically published around the following times of the year — FOMC projections in mid-February and mid-July; Greenbook forecasts in late January, early May, late June, and late October; CBO projections in late January and mid-August; Administration projections in late January and mid-July; Blue Chip forecasts on the tenth day of February, May, July, and November; and SPF forecasts in mid-February, mid-May, mid-August, and mid-November. Forecasts were typically finalized slightly before these publication dates, the lag probably being greatest for the Administration, which usually completed its forecasts in early December and early June. Were we to group forecasts on the basis of finalization date rather than publication date, the Administration’s “winter” forecast might be better characterized as “autumn” and its “summer” forecast as “spring”. (Grouping by quarter, rather than season could have a similar effect). Although that might be appropriate for some purposes, it would complicate our analysis and make little difference to the main results.

[9] To approximate the missing autumn projections for the FOMC, the CBO, and the Administration, we need their estimates of prior-year conditions at the time of their subsequent winter projections. This information is available for the CBO and the Administration but not for the FOMC. For the latter, we use the estimate of prior-year conditions made by the Federal Reserve Board staff and circulated to the Committee in the January Greenbook. Given the similarity of the CBO, Administration, and Greenbook

A second important variation concerns the horizon of the forecast. Historically, the Committee's projections have the shortest horizon, generally covering only the current year in the case of the winter projection and the following year in the case of the summer projection. In contrast, the horizons of the Greenbook, Blue Chip, and SPF projections extend over the following year, and they extend over a third year in the case of the autumn Greenbook forecasts. Finally, the projections published by both the CBO and the Administration extend many years into the future, thereby providing us with information on the accuracy of longer-range projections.

A final variation in data coverage concerns the availability of forecasts of the three main series used in our analysis — real GDP growth, the unemployment rate, and CPI inflation. With the exception of the FOMC, all our sources published projections of these economic measures.[10] In contrast, the Committee published inflation projections based on the total CPI from 1989 through 1999 only; prior to this period, participants based their inflation forecasts on the GNP deflator, and after this period they based them first on the overall PCE price index and later on the core PCE price index. Because these price measures have varying degrees of predictability — in part reflecting differences in their sensitivity to volatile food and energy prices — the FOMC's average historical accuracy in predicting inflation is a mixed estimate, not a "pure" one that can be used to gauge the accuracy of either total PCE or core PCE inflation forecasts. Thus, we do not use the Committee's inflation forecasts in our analysis.

Variations in reporting basis

Our six data sources also differ in the reporting basis of their projections of real activity and inflation. The FOMC, the Greenbook, the Administration, and the Blue Chip all published their projections for real GDP growth, CPI inflation, and the unemployment rate on the same basis now used by the Committee — that is, as fourth-quarter-over-

---

estimates, we suspect that our results are not overly sensitive to any discrepancy between the Greenbook estimates and those made by FOMC participants.

[10] Prior to 1992, all our sources released projections of real GNP instead of real GDP because the former was the measure of real aggregate output featured at the time in the national income accounts. Thus, all references to GDP in this paper should be understood as referring to GNP prior to 1992. In addition, the Administration's unemployment rate forecasts prior to 1992 were for the total unemployment rate (which includes the armed forces), not the civilian unemployment rate projected by our other sources. We have adjusted for this difference in calculating forecast errors.

fourth-quarter percent changes for output growth and inflation and as fourth-quarter averages for the unemployment rate. In contrast, the CBO published projections for real GDP growth and CPI inflation on the desired reporting basis only for the current and following year; it reported projections for more-distant years as calendar-year-over-calendar-year percent changes. Moreover, the CBO did not report projections of the fourth-quarter average of the unemployment rate — only its annual average. Finally, although SPF projections for real activity in the current year are always available on the desired reporting basis, the SPF winter, spring and summer projections for the next year are available only on a calendar-year-over-calendar-year basis for real GDP growth and on an annual-average basis for the unemployment rate.[11] However, SPF forecasts for CPI inflation are always available on the desired Q4-over-Q4 reporting basis for both the current and following years.

These differences in reporting bases create a comparability problem for our analysis, especially in the case of the unemployment rate. Annual unemployment rate projections tend to be more accurate than forecasts of the fourth-quarter average for two reasons. First, averaging across quarters eliminates some quarter-to-quarter noise. Second, the annual average is effectively closer in time to the forecast than the fourth-quarter average because the midpoint of the former precedes the midpoint of the latter by more than four months. This shorter effective horizon is especially important for current-year projections of the unemployment rate because the forecaster will already know, or have a good estimate of, some of the quarterly data that enter the annual average. For this reason, we do not use CBO prediction errors in computing the average accuracy of forecasts of the unemployment rate in the fourth quarter of the current year.

Annual-average forecasts of the unemployment rate probably have a comparative advantage at longer horizons as well for the same reasons. Moreover, similar considerations apply to out-year projections of real GDP growth and CPI inflation made on a calendar-year-over-calendar-year basis. Based on a comparison of Greenbook errors for forecasts made on these different reporting bases, we do not believe that the

[11] The SPF reports forecasts in two ways — as quarterly projections of real GDP, inflation, unemployment, and other variables for the prior quarter, the current quarter, and each of the next four quarters; and as projections of annual averages. By combining the quarterly forecast data with information from the real-time database maintained by the Federal Reserve Bank of Philadelphia, we can construct forecasts of real GDP growth on the desired basis for the current year for all four release dates.

comparability problems are so severe as to merit excluding the out-year CBO and SPF projections from our estimates of average predictive accuracy. However, these reporting differences probably do account for some of the observed (small) differences in forecasting accuracy discussed below.[12]

Defining "truth"

Given our six sources of historical forecast data, we face the issue of how to define "truth" for the purposes of computing prediction errors. One possibility is to use the currently published estimates of historical data for real GDP growth, the unemployment rate, and the CPI. Using currently published data, however, has the drawback of incorporating subsequent definitional changes to the series that forecasters were actually projecting at the time. One example of such a change is the adoption of chain aggregation by the Bureau of Economic Analysis (BEA) in the mid-1990s for constructing measures of real GDP and its components as well as their associated price indexes; another example is the 1999 redefinition of business fixed investment to include outlays for computer software. Using current data would mean that these definitional changes would be a source of forecast error — even though we do not interpret them as a source of uncertainty in a meaningful sense.

To minimize these problems, we define "truth" for real GDP and the two PCE price indexes for each year to be the BEA's so-called "first final" estimate. The "first final" estimate is the third and last one published by the BEA prior to the release of its annual revisions of the national accounts; the BEA usually publishes the first final estimates for the fourth quarter of the prior year in late March. This approach does not entirely free us from the problem of unanticipated methodological revisions for out-year forecasts because some revisions of this sort did occur within two or three years after some of the projections in our sample were made. In the case of the unemployment rate and the total CPI, we use the prior-year estimates reported in the April/May Greenbooks. However, the definition of "truth" is not an important problem for these two series

[12] For example, the root mean squared error of autumn Greenbook forecasts is .62 for next year's fourth-quarter average of the unemployment rate and .42 for the annual average. For two years ahead, the corresponding root mean squared errors are .90 and .76. For real GDP growth and CPI inflation, the differences in accuracy are smaller between calendar-year-over-calendar-year projections and fourth-quarter-over-fourth-quarter projections.

because they are usually subject to only very small revisions relating to estimated seasonal factors.[13]

Mean versus modal forecasts

Another issue important to our forecast comparisons is whether they represent mean predictions as opposed to median or modal forecasts. The projections now produced by FOMC participants are explicitly modal forecasts in that they represent participants' projections of the most likely outcome, with the distribution of risks about the published projections viewed at times as materially skewed.[14] However, we do not know whether participants' projections in the past had this modal characteristic. In contrast, the CBO's forecasts, past and present, are explicitly mean projections. In the case of the Greenbook projections, the Federal Reserve Board staff typically views them as modal forecasts but does not regard the practical difference from a mean projection as usually important. As for our other sources, we have no reason to believe that they are not mean projections, although we cannot rule out the possibility that some of these forecasters may have had some objective other than minimizing the root mean squared error of their predictions. In the case of the SPF and Blue Chip forecasts, the fact that the reported projections are means of many individual forecasts may tend to push them closer to mean predictions even if the underlying projections have modal characteristics.

Implications of conditionality

A final issue of comparability concerns the conditionality of forecasts. Currently, each FOMC participant conditions his or her individual projection on "appropriate monetary policy," defined as the future policy most likely to foster trajectories for output and inflation consistent with the participant's interpretation of the dual mandate. Although the definition of "appropriate monetary policy" was less explicit in the past, Committee participants presumably had a similar idea in mind when making their forecasts historically. Whether or not the other forecasters in our sample (aside from the

---

[13] A possible exception to this statement occurred in 1994 when the Bureau of Labor Statistics made several important changes to the household labor market survey.

[14] This point is illustrated by the discussions of uncertainty in the *Summary of Economic Projections* released November 20, 2007, February 20, 2008, May 21, 2008, and July 15, 2008, in which financial market stress is cited as a factor skewing the risks to the outlook for real activity to the downside.

Greenbook) generated their projections on a similar basis is unknown, but we think it reasonable to assume that most sought to maximize the accuracy of their predictions and so conditioned their forecasts on their assessment of the most likely outcome for monetary policy.

This assumption is not valid for the Greenbook because the Federal Reserve Board staff, in order to avoid inserting itself into the FOMC's internal policy debate, eschews guessing what monetary policy actions would be most consistent with the Committee's objectives. Instead, the staff has traditionally conditioned the outlook on a "neutral" assumption for policy. At times, this approach has taken the form of an unchanged path for the federal funds rate. More typical, however, were paths that modestly rose or fell over time; these trajectories were chosen to signal the staff's assessment that macroeconomic stability would eventually require some adjustment in policy. In principle, these conditioning paths may not have represented the staff's best guess for the future course of monetary policy, and so could have impaired the accuracy of Greenbook projections. Nevertheless, the practical import of this issue seems small, in part because alternative forecasts of interest rates (such as those of the SPF) were no more accurate.[15]

Fiscal policy represents another area where conditioning assumptions could have implications for using historical forecast errors to gauge current uncertainty. The projections reported in the Monetary Policy Report, the Greenbook, the Blue Chip, and the Survey of Professional Forecasters presumably all incorporate assessments of the most likely outcome for federal taxes and government outlays. This assumption is often not valid for the forecasts produced by the CBO and the Administration because the former conditions its baseline forecast on unchanged policy and the latter conditions its baseline projection on the Administration's proposed fiscal initiatives. As was the case with the Greenbook's approach to monetary policy, the practical import of this type of "neutral" conditionality for this study may be small. In particular, such conditionality would not have a large effect on longer-run predictions of aggregate real activity and

[15] The RMSE of 4-quarter-ahead forecasts of the 3-month Treasury bill rate, using all forecasts from 1986 to 2006 was 1.34 percentage points for both the SPF and the Greenbook. For 10-year Treasury bonds, 4-quarter-ahead RMSEs for 1992 (when the SPF series begins) to 2006 were 0.99 percentage points for both the Greenbook and the SPF.

inflation if forecasters project monetary policy to respond endogenously to stabilize the overall macroeconomy. (Of course, fiscal assumptions would matter for forecasts of the budget deficit and perhaps interest rates.)

4. Historical Forecast Accuracy

In this section, we review the empirical evidence on historical predictive accuracy provided by our sample of six forecasters. We organize this review around six key findings, starting first with the general magnitude of forecasting errors and ending with a comparison of our error-based estimates of uncertainty to ones derived from stochastic simulations of the FRB/US model. In all of this analysis, we focus on the root mean squared prediction errors made by our panel.[16]

Result #1: Forecasts errors are large

The difference between actual outcomes, as measured using real-time data, and the forecasts discussed in the previous section represent our set of forecast errors. We calculate these errors for all forecasts published between 1986 and 2006. For each forecaster we then take the square root of the mean squared error, or RMSE. The RMSE is a standard measure of the "typical" forecast error. Figure 1 and Tables 2 through 4 show the average across forecasters of the individual RMSEs for each horizon and variable.[17] We should stress that these average RMSEs are not the root mean squared errors of a hypothetical pooled forecast that someone might have constructed by averaging the forecasts of the different members of our sample. Rather, we average the

---

[16] One topic that we do not address is the average bias of past forecasts, mainly because any previous bias is unlikely to be a useful guide to future prediction errors. This follows because past average errors probably mostly reflect one-time unexpected persistent events — such as the opportunistic disinflation of the late 1980s and early 1990s, the productivity acceleration of the mid-1990s, and the recent extended surge in oil prices — rather than an inherent systematic tendency for forecasters to over- or under-predict some series. Although unexpected persistent events will continue to occur, there is no reason to believe that they will led to forecasting errors of the same sign and magnitude as those experienced in the past. In any event, mean prediction errors in our sample are small from an economic perspective, amounting to only 0.1 or 0.2 percentage point at most. In addition, most (albeit not all) of the mean errors for different combinations of variable and projection horizon are statistically insignificant; bias appears to be more important for the unemployment rate and CPI inflation than it does for real GDP growth.

[17] To construct Figure 1, we allocate each of our four forecast rounds to a quarter – for example, the winter forecast is assumed to be published in the first quarter of the year. This correspondence holds on average across our panel of forecasters, but not for each individual forecast — for example, the Administration's "winter" forecast is constructed in Q4, the FOMC's "summer" forecast is constructed in Q2, and so on.

individual RMSEs of our forecasters in order to generate a benchmark for the typical amount of uncertainty we might expect to be associated with the separate forecasts of the different members of our sample, including the FOMC.

By way of a guide to Figure 1, the shortest forecast horizon we consider is for forecasts published in the autumn — roughly, the fourth quarter of each year — for outcomes of that year, which we label a 0-quarter-ahead error. As shown in the top panel of Figure 1, the average RMSE at this horizon for real GDP growth on a fourth-quarter-over-fourth-quarter basis is 0.6 percentage point.[18] For more distant events, uncertainty is greater. The longest horizon forecasts in our sample are those published in the winter of each year for outcomes that occur three years later in the fourth quarter. At this horizon, the average RMSE for GDP growth is 1.5 percentage points. Similarly, the average RMSE in our sample widens from around a tenth or two to about 1 percentage point for both the unemployment rate and CPI inflation.

These errors seem large. Suppose, for example, that a forecaster projected the unemployment rate to be close to 5 percent over the next few years. Given the size of past errors, we should not be surprised to see the unemployment rate actually climb to 6 percent or fall to 4 percent, because of unanticipated disturbances to the economy and other factors. This fact is sobering because such differences in actual outcomes for the real economy would imply very different states of public well-being and would likely have important implications for the stance of monetary policy. Similarly, an inflation outcome of 1 percent per year would no doubt be seen by the FOMC as having quite different ramifications for the appropriate level of the federal funds rate from an inflation outcome of 3 percent. Yet, we should not be overly surprised to see either inflation outcome if we are projecting prices to rise 2 percent per year.

Another way of gauging the size of these forecast errors is to compare them to the actual variations seen from year to year over history. We might judge forecasters as informative if they make errors that on average are only a fraction of the standard deviations for output growth, the unemployment rate, and inflation. However, the ratio of the average RMSEs to the 1986-2006 standard deviations is about 90 percent for real

[18] This RMSE takes account of errors in forecasts of quarterly real GDP growth in both the third and fourth quarters because the first final estimate of real GDP for the third quarter is not available until late December.

GDP growth, 50 percent for the level of the unemployment rate, and 80 percent for CPI inflation in the case of winter projections of current-year conditions. For longer forecast horizons, RMSEs are about the same as the standard deviations. Thus, GDP and inflation forecasts explain very little of the future variation in these series on average; alternatively put, sample means are about as accurate a guide to the future as published forecasts, at least beyond the very near term. This striking result has been documented for the SPF (Campbell, 2007), the Greenbook (Tulip, 2005), and other large industrial economies (Vogel, 2007). It has important implications for forecasting and policy which go beyond the scope of this paper.

As Figure 1 illustrates, average RMSEs increase with forecast horizon and then tend to flatten out. This result is not surprising because we know more about the forces affecting near-term events; put another way, as the time between a forecast and an event increases, more surprises will accumulate. That said, much of the widening of the RMSEs for GDP growth and inflation reflects data construction rather than increasing uncertainty about the future. Near-horizon forecasts of real GDP growth and CPI inflation span some quarters for which the forecaster already has published quarterly data. For this reason, most of the increase in RMSEs during the first few quarters essentially reflects the incremental replacement of a known past with an unknown future in the calculations.

Result #2: Differences across forecasters are small

Our second main result is that differences in predictive accuracy across forecasters are small. This point is evident from a simple perusal of Tables 2 through 4, which shows that RMSEs on the same reporting basis and for the same variable-horizon combination typically differ by only one or two tenths of a percentage point across forecasters, controlling for release date. Compared with the size of the RMSEs themselves, such differences seem relatively unimportant because they imply only modest variations in the average magnitude of past uncertainty. Moreover, some of the differences clearly reflect the variations in reporting basis that we discussed earlier. For example, as was discussed in Section 3, the CBO’s unemployment forecasts were more accurate than others were in part because they are on an annual-average basis. Finally, some of variation across forecasters likely reflects differences in the timing of projections because forecasts made late in the quarter tend to be more accurate than those made earlier, particularly in the case of current-year forecasts.[19]

Of course, some of the differences we observe probably reflect random noise, given the small size of our sample. Table 5 shows p-values from a test of the hypothesis that the RMSEs are unequal because of chance alone — that is, the probability that we would see such differences because of random sampling variability when all forecasters are in fact equally accurate.[20] Generalizing about the results in the table is difficult, given that the tests are not independent. Nevertheless, the broad pattern is for p-values to be large for the longer-horizon forecasts — that is, the likelihood is high that accuracy is the same for out-year forecasts. In contrast, some clear differences do appear for current-year and next-year projections. However, we judge that timing and methodological differences probably account for most of these low p-values, in part because the low next-year p-values tend to become insignificant when we exclude the projections reported on a non-standard basis (that is, some of the CBO and SPF projections).

That the forecasts in our sample have similar accuracy is perhaps not particularly surprising, given that each reflects the average view of many people, either explicitly or implicitly. This similarity is most clear for the Blue Chip and the SPF forecasts, which represent the mean of the individual projections provided by a large group of forecasters. By averaging across many independent projections, these surveys tend to wash out idiosyncratic differences in forecasting techniques or views about the economy. Given that the participants in both surveys have similar expertise and the same access to information, we therefore should not expect these two surveys to yield appreciably different views about the outlook. A similar logic applies to the midpoint of the

---

[19] Although it is difficult to characterize the typical order in which our six forecasters released their projections (as the order varied over time), we judge that the Administration's projections were the earliest on average and that the SPF projections were the latest. Some calculations on our part suggest that timing differences might explain as much as 0.1 percentage point of the current-year differences in accuracy but considerably less at longer horizons.

[20] The test statistic is a generalization of the Diebold and Mariano (1995) test of predictive accuracy. In comparing two forecasts, one implements the test by regressing the difference between the squared errors for each forecast on a constant. The test statistic is a *t*-test of the hypothesis that the constant is significantly different from zero once allowance is made for the errors having a moving average structure. For comparing *n* forecasts, we construct *n-1* differences and jointly regress these on *n*-1 constants. The test statistic that these constants jointly equal zero is asymptotically chi-squared with *n*-1 degrees of freedom, where again allowance is made for the errors following a moving average process.

consensus range of the FOMC projections published in the *Monetary Policy Report*, as this forecast is the "average" view of nineteen informed individuals. Of course, our three other sources — the Federal Reserve Board staff, the Congressional Budget Office, and the Administration — do produce forecasts that represent the views of individual institutions. But in all three cases the forecasting process incorporates the analysis and judgment of a large number of economists, and for all practical purposes the outcome is a group decision.[21] Furthermore, these institutions are aware of the private forecasts when they make their own forecasts, and knowing that the Blue Chip and the SPF often contain valuable information, may adjust their projections in ways that bring them into closer alignment with the consensus of private forecasters.

The forecasters in our sample not only display a similar degree of accuracy on average, they also tend to make similar individual prediction errors over time — a phenomenon that both Gavin and Mandal (2001) and Sims (2002) have noted. This tendency reveals itself in correlations between the errors made by different forecasters that typically range from 0.75 and 0.95. That forecasters make similar mistakes does not seem surprising. All forecasters use the past as a guide to the future, and so any deviation from average historical behavior in the way the economy responds to a shock will tend to result in common projection errors. Moreover, such apparent deviations from past behavior are not rare, both because our understanding of the economy and its evolution is limited, and because shocks rarely (if ever) repeat themselves in all their particulars. Finally, some economic disturbances are probably inherently difficult to predict in advance, abstracting from whether or not forecasters clearly understand their economic consequences once they occur. Based on these considerations, it is not surprising that highly correlated prediction errors would result from such events as the financial "headwinds" period of the early 1990s, the acceleration in productivity that began in the mid-1990s, the expansion and bursting of the dot-com bubble, and the financial strains that have accompanied the collapse in subprime mortgage lending.

---

[21] The Administration's forecast is actually the joint product of three agencies: the Council of Economic Advisers, the Treasury Department, and the Office of Management and Budget. Representatives from the three groups (known as the "Troika") participate at all stages of the forecasting process, and the projection is not official until the head of each agency formally signs off.

Although the focus of this paper is on the performance of forecasters as a whole and not that of any individual forecaster, we think one particular result is worthy of note. Romer and Romer (2000) and Sims (2002) reported that the Greenbook, over a period that extended back into the 1970s and ended in the early 1990s, outperformed some other forecasters, especially for short-horizon forecasts of inflation. In contrast, a review of Tables 2 through 4 reveals that the Federal Reserve Board staff performs about the same as other forecasters for our sample.

Result #3: RMSE-based bands span roughly 70 percent of errors

If we knew that forecast errors were normally distributed, then we would expect to see roughly 70 percent of historical errors to fall within a range equal to zero plus or minus one RMSE. (The precise theoretical probability is 68 percent.) This knowledge, in turn, would enable forecasters to make probability statements of the sort we noted in the introduction: "Based on the historical accuracy of our forecasts, we believe there is about a 70 percent probability that variable X will lie within plus or minus one RMSE of the current projection." However, forecasters could still make statements of this sort without knowledge of the exact distribution (whether normal or something else) as long as they had a good sense of the percentage of past errors that did fall within a range of plus or minus one RMSE.

Table 6 shows the actual percentage of past errors — for each combination of variable, forecast horizon, and release date — that did fall within this range. Most of the percentages are in the vicinity of 68 percent. Of course, the actual frequency of errors falling in a given band will vary in small samples, reflecting random variation. If the true percentage is 68 percent and our sample includes N independent errors, the actual percentages we observe should be binomially distributed with parameters (N, .68). Accordingly, we can compute the probability of observing deviations from 68 percent that are as large as we see. As indicated by the asterisks in Table 6, this probability is less than 10 percent in only a few cases using a two-tailed test. These results seem consistent with our assumptions, especially given that our errors are not independent (thereby implying that our test overstates the effective size of our sample).

For some purposes, such as statistical inference or the calculation of confidence intervals with different probabilities, we would like to know if the forecast errors are in fact distributed normally; for example, we could then say that 95 percent of the errors should fall within a range equal to plus or minus two RMSEs. Figures 2 and 3 provide some graphical information on the historical distribution of forecast errors for the winter and spring publication dates, respectively; the figures also compare these empirical histograms to a corresponding theoretical normal distribution with mean zero and standard deviation equal to the sample RMSE. To construct these graphs for the different publication dates, we pool the prediction errors for all the forecasters in our sample for each different variable-horizon combination. Visually, the actual empirical distributions overall do not appear to be grossly at odds with the theoretical normal distribution. For example, the empirical and theoretical distributions appear to have roughly the same amount of mass concentrated in the central shaded region, which encompasses 68 percent of the mass under the normal distribution.

Overall impressions aside, some of the histograms do seem to have suspiciously non-normal features, such as skewness. Jarque-Bera tests for normality yield p-values of 5 percent or less for almost one-third of all the different combinations of variable, forecast horizon, and release date (Table 7). By themselves, these results would seem to cast doubt on the validity of assuming that forecast errors are normally distributed. However, the Jarque-Bera test in these circumstances is likely to signal non-normality too often because the errors are highly correlated across forecasters; forecast errors for horizons beyond the current year are serially correlated as well.

If we were less comfortable with the assumption of normality, an alternative approach to confidence interval estimation would be to use the 15^{th} and 85^{th} percentiles of the actual distributions. For a sample of 20, that is the range between the 4^{th} and 17^{th} (inclusive) ranked observations. We prefer to use root mean squared errors, partly for their familiarity and comparability with other research, and partly because they are probably less sensitive to small-sample quirks in the data. However, if we were to take this alternative approach, the practical effect would be to increase the width of the 70-percent confidence intervals for real GDP about 6 percent on average. Such a shift would imply a rise in the estimated interval for, say, summer-release forecasts of third year GDP growth from 1.84 percentage points under our preferred approach to 2 percentage points under the alternative method. The corresponding shift in the estimated confidence intervals for the unemployment rate would be somewhat larger at about 11 percent on average, while the average shift for inflation intervals would be close to zero.

Result #4: CPI inflation errors may be larger than PCE inflation errors

The FOMC presents inflation forecasts for the total PCE price index and the core PCE price index. Given the scarcity of historical forecasts of these variables, we derive a benchmark for past inflation uncertainty from forecasts for total CPI inflation. The CPI is closely related to the PCE price indexes (indeed, many components of the PCE measure are derived directly from detailed pieces of the CPI) and has been widely forecast. However, there are important differences as well, including the weights assigned to different elements of the price indexes, the scope of expenditures considered, and the methods used to aggregate detailed components into the overall measure (fixed weights versus chain weighting).

Figure 4 compares forecast errors for these three series based on Greenbook projections. Over most of the forecast horizon, RMSEs for PCE inflation are about 10 percent smaller than RMSEs for CPI inflation. Core PCE RMSEs are as much as 38 percent smaller than CPI RMSEs (for four-quarter-ahead forecasts), although the margin falls to 25 percent for nine-quarter-ahead forecasts. (These latter forecasts appeared in September Greenbooks and were for inflation two calendar years ahead.) These comparisons suggest that the errors we report for CPI inflation slightly overstate past uncertainty about PCE inflation and considerably overstate past uncertainty for core PCE inflation.

Result #5: Error-based estimates of uncertainty are sensitive to sample period

Ideally, our RMSE estimates of forecast uncertainty should be reasonably robust, in the sense that they should not be markedly sensitive to modest adjustments to our basic methodology, such as changes in sample period. In Table 8, we report average RMSEs for three different samples — the full sample covering the period from 1986 to 2006, a truncated sample that excludes forecasts made prior to 1991, and a third sample that excludes all projections made after 2000. By dropping projections made at either the start or the end of the full sample period, we cause the RMSEs to shift by as much as two or three tenths of a percentage point. Such shifts are not huge from an economic standpoint, but they do illustrate how relatively minor variations in sample period can cause estimated confidence intervals to shift by a noticeable amount.

Estimates of historical forecasting accuracy look much less stable if one examines the period prior to the start of the Great Moderation in the mid-1980s. Campbell (2007) for the SPF and Tulip (2005) for the Greenbook find statistically and economically significant reductions in the size of forecast errors in the mid-1980s. For example, Tulip reports that the root mean squared error of the Greenbook forecast of real GDP growth was roughly 40 percent smaller after 1984 than before, while the RMSE for the GDP deflator fell by between a half and two-thirds. These changes highlight a general risk to any assessment of uncertainty: The economy can evolve over time in ways that can fundamentally alter its inherent predictability. The fact that the economy has been unusually calm for the last twenty years is not a guarantee that it will remain so permanently.

Result #6: Error-based and model-based estimates of uncertainty are broadly similar

Our final empirical result concerns the relationship between our RMSE-based estimates of uncertainty and ones derived from simulations of econometric models.[22] Model-based estimates, like error-based estimates, are by nature empirical and rooted in historical experience, in part because the parameters of an econometric model are typically chosen to fit the data subject to the constraints imposed by economic theory. In addition, the model-based approach attempts to approximate the probability distribution of possible outcomes for the future by repeatedly simulating the response of the model to random shocks of the sort seen in the past. Despite this shared empirical basis, however, model-based estimates may differ noticeably from error-based estimates for many reasons. As was noted earlier, the model-based approach may understate uncertainty because after-the-fact information was used to design and estimate the model;

[22] For an early example of this approach using a macroeconometric model of the United States, see Fair (1980). Garratt, Lee, Pesaran and Shin (2003) provide a more recent example for the United Kingdom.

alternatively, it may overstate uncertainty by ignoring extra-model information available to forecasters at the time.

With these caveats in mind, Table 9 compares the average RMSEs from our sample with forecast standard error estimates derived from simulations of FRB/US, an econometric model used by the Federal Reserve Board staff for forecasting and policy analysis.[23] FRB/US-based estimates are particularly interesting in the present context because the staff uses them to construct the fan charts and other uncertainty estimates that have been a regular feature of the Greenbook in the past few years. To construct these model-based estimates, we simulate the model 5000 times, repeatedly subjecting it to shocks drawn randomly from the 1986-2006 set of model equation residuals. In the simulations, monetary policy responds to the shocks as predicted by an estimated version of the Taylor rule.

As can be seen, the model-based estimates are remarkably similar to the historical RMSEs for both real GDP and inflation. In contrast, the model results suggest that uncertainty about the outlook for the unemployment rate is markedly less than our estimates of past forecasting performance would suggest.[24] Table 9 also provides some supplementary information on the relative magnitude of uncertainty estimates for total CPI inflation, total PCE inflation, and core PCE inflation. According to the FRB/US model, uncertainty about the outlook for total PCE inflation appears to be somewhat less than uncertainty about total CPI inflation, while uncertainty about core PCE inflation turns out to be only two-thirds as large as total CPI uncertainty. These relative magnitudes are roughly consistent with the evidence presented earlier using historical Greenbook RMSEs.

5. Conclusions

In this paper, we have presented estimates of past forecast uncertainty; these estimates are now used by FOMC participants as a benchmark against which to assess the

[23] An overview of the FRB/US model is in Brayton and Tinsley (1996). Additional information about the model is provided by Brayton, Mauskopf, Reifschneider, Tinsley, and Williams (1997); Brayton, Levin, Tryon, and Williams (1997); and Reifschneider, Tetlow, and Williams (1999). An example of the use of FRB/US simulations to estimate uncertainty is Reifschneider and Williams (2000).

[24] We are inclined to discount this particular finding because it may be an artifact of the design of the model; for example, certain labor market factors that contributed to the historical variation in unemployment are effectively exogenous in the FRB/US simulations.

uncertainty of the current economic outlook. This approach, which exploits the historical forecast record of several groups, suggests that uncertainty about the economic outlook is considerable — a point emphasized by the Federal Open Market Committee in their communications on this issue.

We should repeat a caveat to our analysis: Our approach rests to a large degree on the assumption that the past is a good guide to the future and that forecasters in the future will make prediction errors similar to those made over the past twenty years. Although assumptions of this sort are a practical necessity in all empirical work, we must bear in mind that the economy appears to have been less predictable before the mid-1980s. Because forecast accuracy has changed in the past, it could change again, for better or worse. If so, our error-based estimates by themselves would paint a misleading picture of the potential risks to the outlook.

For this and other reasons, we stress that our approach is not a complete assessment of uncertainty at present. Rather, our estimates should be used in conjunction with the qualitative summary the FOMC publishes of participants' views concerning the uncertainty of the current outlook. Recently, for example, Committee participants have suggested that uncertainty is greater than that seen on average in the past for both real activity and inflation; many participants also view the risks to the outlook as skewed to the downside for output and to the upside for inflation. This information, taken as a whole, conveys a more complete sense of likely outcomes than would be provided by their individual forecasts and the benchmark uncertainty estimates alone.

References

Bernanke, Ben S. (2007). "Federal Reserve Communications," speech delivered at the Cato Institute 25th Annual Monetary Conference, Washington (November 14).

Brayton, Flint, and Peter Tinsley (1996). "A Guide to FRB/US: A Macroeconomic Model of the United States," Financial and Economics Discussion Series 1996-42, Washington: Federal Reserve Board (October).

Brayton, Flint, Eileen Mauskopf, David Reifschneider, Peter Tinsley, and John C. Williams (1997). "The Role of Expectations in the FRB/US Macroeconomic Model," *Federal Reserve Bulletin*, 83 (April), 227-245.

Brayton, Flint, Andrew Levin, Ralph Tryon, and John C. Williams (1997). "The Evolution of Macro Models at the Federal Reserve Board," *Carnegie-Rochester Conference Series on Public Policy*, 4 (December), 43-81.

Campbell, Sean D. (2007). "Macroeconomic Volatility, Predictability, and Uncertainty in the Great Moderation: Evidence from the Survey of Professional Forecasters," *Journal of Business and Economic Statistics*, 25 (April), 191-200.

Diebold, Francis X., and Roberto S. Mariano (1995). "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics*, 13 (July), 253-263.

Fair, R. C. (1980): "Estimating the Expected Predictive Accuracy of Econometric Models," *International Economic Review*, 21, 355-378.

Federal Open Market Committee (2007). "FOMC Announces It Will Increase the Frequency and Expand the Content of Economic Projections Released to the Public," press release, November 14.

Garratt, Anthony, Kevin Lee, M. Hashem Pesaran, and Yongcheoi Shin (2003): "Forecast Uncertainties in Macroeconomic Modeling: An Application to the U.K. Economy," *Journal of the American Statistical Association*, 98, 829-838.

Gavin, William T. and Rachel J. Mandal (2001): "Forecasting Inflation and Growth: Do Private Forecasts Match Those of Policymakers?" *Business Economics*, January, 13-20.

McConnell, Margaret M., and Gabriel Perez-Quiros (2000). "Output Fluctuations in the United States: What Has Changed Since the Early 1980s?" *American Economic Review*, 90 (December), 1464-1476.

Mishkin, Frederic S. (2007). "Inflation Dynamics," speech delivered at the Annual Macro Conference, Federal Reserve Bank of San Francisco, San Francisco, Calif. (March 23).

Reifschneider, David, Robert Tetlow, and John C. Williams (1999). “Aggregate Disturbances, Monetary Policy, and the Macroeconomy: The FRB/US Perspective,” *Federal Reserve Bulletin*, 85 (January), 1-19.

Romer, Christina D., and David H. Romer (2000). “Federal Reserve Information and the Behavior of Interest Rates,” *American Economic Review*, 90 (June), 429-457.

Sims, Christopher A. (2002). “The Role of Models and Probabilities in the Monetary Policy Process,” *Brookings Papers on Economic Activity*, 2002 (2), 1-62.

Stock, James H., and Mark W. Watson (2007). “Why Has U.S. Inflation Become Harder to Forecast?” *Journal of Money, Credit and Banking*, 39 (February, Supplement 1), 3-33.

Tay, Anthony S. and Kenneth F. Wallis (2000): “Density Forecasting: A Survey,” *Journal of Forecasting*, 19, 235-254.

Tulip, Peter (2005). “Has Output Become More Predictable? Changes in Greenbook Forecast Accuracy,” Finance and Economics Discussion Series 2005-31, Washington: Federal Reserve Board (August).

Vogel, Lukas (2007), “How Do the OECD Growth Projections for the G7 Economies Perform? A Post-Mortem”, OECD Economics Department Working Paper, No 573, Paris

Table 1

Variations in Data Coverage and Reporting Basis Across Forecasters

Source	Release Dates	Horizon	Reporting Basis		
			Real GDP Growth	Unemployment Rate	Total CPI Inflation
Monetary Policy Report	Winter Summer	Extends to next year in summer	Q4/Q4	Q4	Not used
Federal Reserve staff (Greenbook)	Winter Spring Summer Fall	Extends to year after next in autumn	Q4/Q4	Q4	Q4/Q4
Congressional Budget Office	Winter Summer	More than four years	Q4/Q4 for current and next year; annual average thereafter	Annual average	Q4/Q4 for current and next year; annual average thereafter
Administration	Winter Summer	More than four years	Q4/Q4	Q4 Total through 1991, then civilian	Q4/Q4
Blue Chip	Winter Spring Summer Fall	Current and next year	Q4/Q4	Q4	Q4/Q4
Survey of Professional Forecasters	Winter Spring Summer Fall	Current and next year	Q4/Q4 for current year and (for autumn only) next year. Otherwise, annual average	Q4 for current year and, for autumn only, next year. Otherwise, annual average	Q4/Q4

Table 2
Root Mean Squared Prediction Errors for Real GDP[1]
(projections published from 1986 to 2006)

	Projection Period Year			
	Current	Second	Third	Fourth
Winter projections				
Monetary Policy Report	1.10	—	—	—
Federal Reserve staff (Greenbook)	1.19	1.39	—	—
Congressional Budget Office	1.26	1.39	1.35[3]	1.43[3]
Administration	1.29	1.48	1.46	1.52
Blue Chip	1.17	1.30	—	—
Survey of Professional Forecasters	1.14	1.31[3]	—	—
Average	1.19	1.37	1.40	1.48
Spring projections				
Monetary Policy Report[2]	0.94	—	—	—
Federal Reserve staff (Greenbook)	1.03	1.25	—	—
Congressional Budget Office[2]	1.03	1.36	1.36[3]	1.40[3]
Administration[2]	1.09	1.44	1.45	1.50
Blue Chip	1.04	1.31	—	—
Survey of Professional Forecasters	0.92	1.31[3]	—	—
Average	1.01	1.33	1.41	1.45
Summer projections				
Monetary Policy Report	0.84	1.28	—	—
Federal Reserve staff (Greenbook)	0.89	1.34	—	—
Congressional Budget Office	0.89	1.35	1.39[3]	1.38[3]
Administration	0.96	1.42	1.45	1.47
Blue Chip	0.90	1.29	—	—
Survey of Professional Forecasters	0.79	1.19[3]	—	—
Average	0.88	1.31	1.42	1.43
Autumn projections				
Monetary Policy Report[2]	0.46	1.16	—	—
Federal Reserve staff (Greenbook)	0.60	1.37	1.50	—
Congressional Budget Office[2]	0.54	1.27	1.35[3]	1.37[3]
Administration[2]	0.61	1.28	1.48	1.47
Blue Chip	0.64	1.26	—	—
Survey of Professional Forecasters	0.61	1.25	—	—
Average	0.58	1.27	1.44	1.42

1. Percent change, fourth quarter of year from fourth quarter of previous year.
2. Average of summer projection and subsequent winter projection.
3. Percent change, annual average for year relative to annual average of previous year.

Table 3
Root Mean Squared Prediction Errors for the Unemployment Rate[1]
(projections published from 1986 to 2006)

	Projection Period Year			
	Current	Second	Third	Fourth
Winter projections				
Monetary Policy Report	0.48	—	—	—
Federal Reserve staff (Greenbook)	0.46	0.79	—	—
Congressional Budget Office	0.33[3]	0.78[3]	1.04[3]	1.07[3]
Administration	0.51	0.85	1.02	1.04
Blue Chip	0.49	0.89	—	—
Survey of Professional Forecasters	0.47	0.71[3]	—	—
Average	0.48[4]	0.80	1.03	1.06
Spring projections				
Monetary Policy Report[2]	0.36	—	—	—
Federal Reserve staff (Greenbook)	0.35	0.73	—	—
Congressional Budget Office[2]	0.21[3]	0.69[3]	1.00[3]	1.09[3]
Administration[2]	0.41	0.80	1.00	1.06
Blue Chip	0.39	0.85	—	—
Survey of Professional Forecasters	0.33	0.63[3]	—	—
Average	0.37[4]	0.74	1.00	1.07
Summer projections				
Monetary Policy Report	0.28	0.77	—	—
Federal Reserve staff (Greenbook)	0.30	0.73	—	—
Congressional Budget Office	0.14[3]	0.66[3]	0.99[3]	1.12[3]
Administration	0.32	0.77	0.99	1.09
Blue Chip	0.29	0.79	—	—
Survey of Professional Forecasters	0.25	0.52[3]	—	—
Average	0.29[4]	0.71	0.99	1.10
Autumn projections				
Monetary Policy Report[2]	0.14	0.60	—	—
Federal Reserve staff (Greenbook)	0.15	0.62	0.90	—
Congressional Budget Office[2]	0.07[3]	0.46[3]	0.87[3]	1.07[3]
Administration[2]	0.20	0.62	0.91	1.05
Blue Chip	0.14	0.65	—	—
Survey of Professional Forecasters	0.11	0.60	—	—
Average	0.15[4]	0.59	0.89	1.06

1. Fourth-quarter average.
2. Average of summer projection and subsequent winter projection.
3. Annual average. Forecasts for the annual average of the unemployment rate tend to be more accurate than forecasts for the fourth-quarter average, especially for the current year.
4. Excludes CBO root mean squared prediction errors because of non-comparability.

Table 4
Root Mean Squared Prediction Errors for the Consumer Price Index[1]
(projections published from 1986 to 2006)

	Projection Period Year			
	Current	Second	Third	Fourth
Winter projections				
Federal Reserve staff (Greenbook)	1.00	0.98	—	—
Congressional Budget Office	0.99	0.90	0.80[3]	0.87[3]
Administration	1.07	1.05	1.04	1.16
Blue Chip	0.95	0.91	—	—
Survey of Professional Forecasters	0.93	0.98	—	—
Average	0.99	0.96	0.92	1.01
Spring projections				
Federal Reserve staff (Greenbook)	0.71	1.02	—	—
Congressional Budget Office[2]	0.73	0.92	0.80[3]	0.88[3]
Administration[2]	0.74	0.98	1.04	1.11
Blue Chip	0.64	0.92	—	—
Survey of Professional Forecasters	0.66	1.02	—	—
Average	0.69	0.97	0.92	0.99
Summer projections				
Federal Reserve staff (Greenbook)	0.52	1.03	—	—
Congressional Budget Office	0.62	0.96	0.81[3]	0.89[3]
Administration	0.62	0.94	1.08	1.08
Blue Chip	0.59	0.94	—	—
Survey of Professional Forecasters	0.50	1.03	—	—
Average	0.57	0.98	0.95	0.99
Autumn projections				
Federal Reserve staff (Greenbook)	0.15	1.07	1.04	—
Congressional Budget Office[2]	0.36	0.91	0.80[3]	0.84[3]
Administration[2]	0.35	0.93	1.07	1.06
Blue Chip	0.21	0.94	—	—
Survey of Professional Forecasters	0.24	0.95	—	—
Average	0.26	0.96	0.97	0.95

1. Percent change, fourth quarter of year from fourth quarter of previous year.
2. Average of summer projection and subsequent winter projection.
3. Percent change, annual average for year relative to annual average of previous year.

Table 5
P-Values from Test of Hypothesis That All Forecasters Have the Same Predictive Accuracy[1]

	Projection Period Year			
	Current	Second	Third	Fourth
Winter projections				
Real GDP	.10	.54	.52	.55
Unemployment Rate	.44[2]	<.01	.78	.77
Total CPI	.05	.22	.25	.28
Spring projections				
Real GDP	.20	.75	.59	.51
Unemployment Rate	<.01[2]	<.01	.87	.78
Total CPI	.48	.06	.22	.31
Summer projections				
Real GDP	.13	.17	.68	.51
Unemployment Rate	.01[2]	<.01	.96	.81
Total CPI	.23	.02	.20	.32
Autumn projections				
Real GDP	.26	.52	.47	.52
Unemployment Rate	<.01[2]	<.01	.53	.83
Total CPI	<.01	.44	.20	.27

1. Multivariate generalization of the Diebold and Mariano (1995) test of predictive accuracy. Details are in footnote 20.
2. Excludes CBO annual-average unemployment rate forecasts.

Table 6
Percent of Historical Errors Falling Within a Band
Equal to Plus or Minus the Average Root Mean Squared Prediction Error[1]
(asterisk denotes significant difference from 68 percent at the 10 percent level)

	Projection Period Year			
	Current	Second	Third	Fourth
Winter projections				
Real GDP (Q4-over-Q4 percent change)	62	66	71	67
Unemployment rate (Q4 average) [2]	66	61	63	67
Total CPI (Q4-over-Q4 percent change)	71	68	66	75
Spring projections				
Real GDP (Q4-over-Q4 percent change)	61*	67	68	67
Unemployment rate (Q4 average) [2]	69	61	61	64
Total CPI (Q4-over-Q4 percent change)	67	70	66	72
Summer projections				
Real GDP (Q4-over-Q4 percent change)	61*	67	74	64
Unemployment rate (Q4 average) [2]	69	65	58	64
Total CPI (Q4-over-Q4 percent change)	70	68	66	69
Autumn projections				
Real GDP (Q4-over-Q4 percent change)	66	68	70	64
Unemployment rate (Q4 average) [2]	69	59*	56*	61
Total CPI (Q4-over-Q4 percent change)	76*	70	70	69

1. Sample includes prediction errors for all forecasters with projections published from 1986 to 2006.
2. Current-year count excludes CBO annual-average unemployment rate forecasts.

Table 7
P-Values from Jarque-Bera Test of Normality of Historical Forecast Errors[1]

	Projection Period Year			
	Current	Second	Third	Fourth
Winter projections				
Real GDP	.17	.59	.80	.35
Unemployment Rate	.08[2]	.01	.21	.52
Total CPI	.32	<.01	.11	.02
Spring projections				
Real GDP	.08	.51	.83	.41
Unemployment Rate	.04[2]	<.01	.21	.45
Total CPI	<.01	<.01	.06	.06
Summer projections				
Real GDP	.07	.41	.86	.48
Unemployment Rate	.01[2]	<.01	.21	.39
Total CPI	<.01	<.01	.04	.18
Autumn projections				
Real GDP	.26	.47	.80	.62
Unemployment Rate	<.01[2]	<.01	.09	.36
Total CPI	<.01	<.01	.01	.12

1. Sample includes prediction errors for all forecasters with projections published from 1986 to 2006.
2. Excludes CBO annual-average unemployment rate forecasts.

Table 8
Sensitivity of Root Mean Squared Prediction Errors to Changes in the Sample Period

	Projection Period Year			
	Current	Second	Third	Fourth
Winter projections				
Real GDP				
1986 to 2006 sample	1.19	1.37	1.40	1.48
1991 to 2006 sample	1.21	1.34	1.24	1.28
1986 to 2000 sample	1.27	1.55	1.56	1.59
Unemployment rate				
1986 to 2006 sample	0.48[1]	0.80	1.03	1.06
1991 to 2006 sample	0.48[1]	0.78	0.94	0.95
1986 to 2000 sample	0.49[1]	0.85	1.09	1.14
Total CPI				
1986 to 2006 sample	0.99	0.96	0.92	1.01
1991 to 2006 sample	0.83	0.94	0.83	0.85
1986 to 2000 sample	0.95	0.87	0.92	1.05
Summer projections				
Real GDP				
1986 to 2006 sample	0.88	1.31	1.42	1.43
1991 to 2006 sample	0.88	1.33	1.28	1.28
1986 to 2000 sample	0.88	1.49	1.58	1.54
Unemployment rate				
1986 to 2006 sample	0.29[1]	0.71	0.99	1.10
1991 to 2006 sample	0.30[1]	0.70	0.92	1.03
1986 to 2000 sample	0.27[1]	0.76	1.06	1.18
Total CPI				
1986 to 2006 sample	0.57	0.98	0.95	0.99
1991 to 2006 sample	0.53	0.94	0.89	0.87
1986 to 2000 sample	0.48	0.91	0.94	1.02

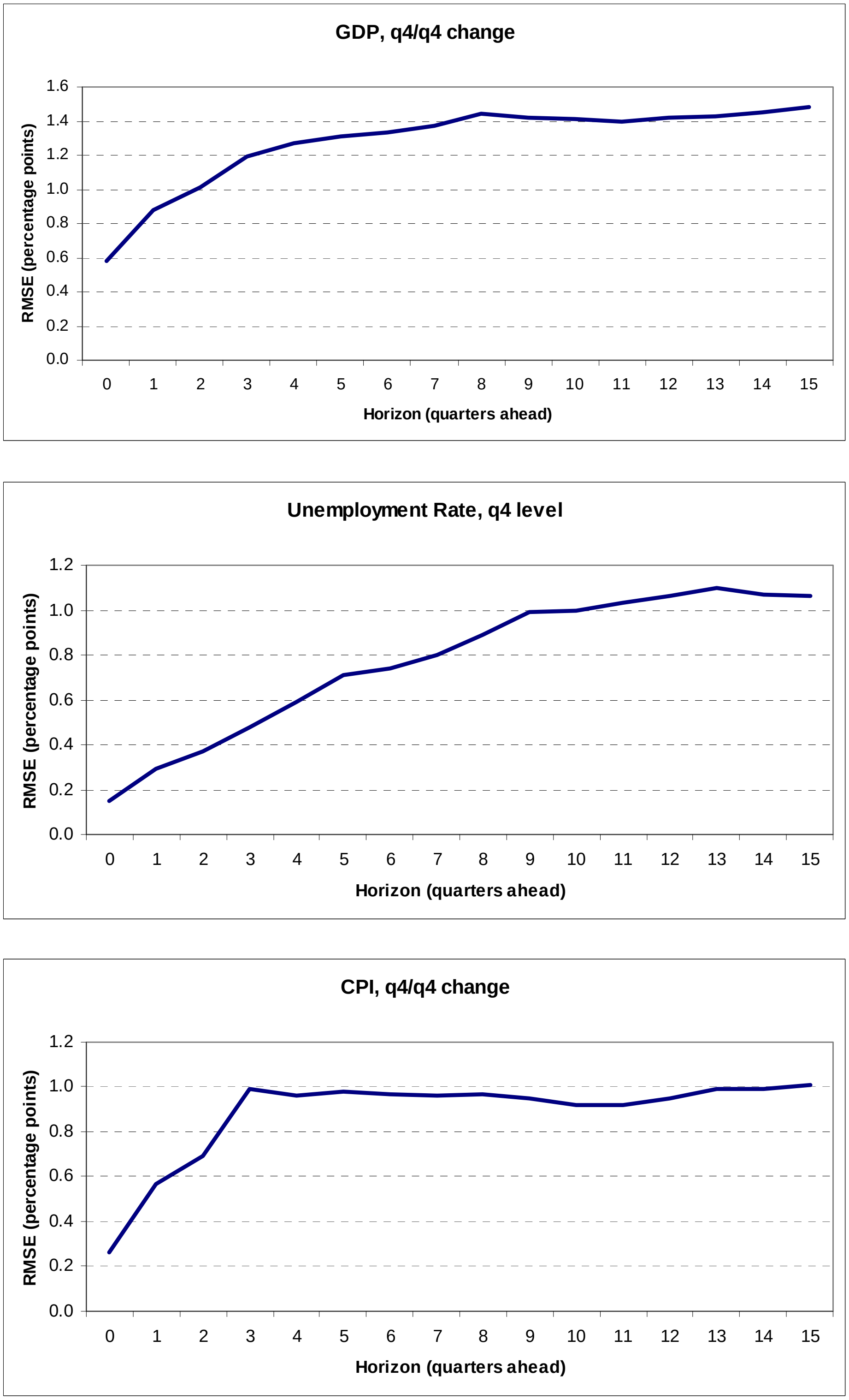
1. Excludes CBO annual-average unemployment rate forecasts.

Table 9
Comparison of Error-Based and Model-Based Estimates of Uncertainty[1]

	Projection Period Year			
	Current	Second	Third	Fourth
Winter projections				
Real GDP (Q4-over-Q4 percent change)				
Average historical root mean squared error	1.19	1.37	1.40	1.48
Model estimate of forecast standard error	1.13	1.35	1.61	1.64
Unemployment rate (Q4 average)				
Average historical root mean squared error	0.48[2]	0.80	1.03	1.06
Model estimate of forecast standard error	0.32	0.59	0.70	0.78
Total CPI (Q4-over-Q4 percent change)				
Average historical root mean squared error	0.99	0.96	0.92	1.01
Model estimate of forecast standard error	0.94	1.00	1.07	1.11
Summer projections				
Real GDP (Q4-over-Q4 percent change)				
Average historical root mean squared error	0.88	1.31	1.42	1.43
Model estimate of forecast standard error	0.81	1.29	1.54	1.58
Unemployment rate (Q4 average)				
Average historical root mean squared error	0.29[2]	0.71	0.99	1.10
Model estimate of forecast standard error	0.20	0.46	0.61	0.74
Total CPI (Q4-over-Q4 percent change)				
Average historical root mean squared error	0.57	0.98	0.95	0.99
Model estimate of forecast standard error	0.63	0.96	1.05	1.08
Addenda: model estimates of PCE price forecast standard errors				
Total (Q4-over-Q4 percent change)				
Winter	0.75	0.84	0.91	0.95
Summer	0.50	0.80	0.88	0.92
Core (Q4-over-Q4 percent change)				
Winter	0.44	0.62	0.71	0.76
Summer	0.27	0.56	0.68	0.72

1. Model estimates based on 5000 stochastic simulations of the FRB/US model, with shocks drawn from the 1986-2006 set of model equation residuals.
2. Excludes CBO annual-average unemployment rate forecasts.

Figure 1:
Average Root Mean Squared Forecast Errors



Note: The forecast horizon extends from the forecast published in the autumn of current year conditions, to the forecast published in the winter of developments fifteen quarters ahead.

Figure 2
Winter Projection Errors -- Actual Distribution Vs. Normal Density
(shaded area equals zero plus or minus average root mean squared error)

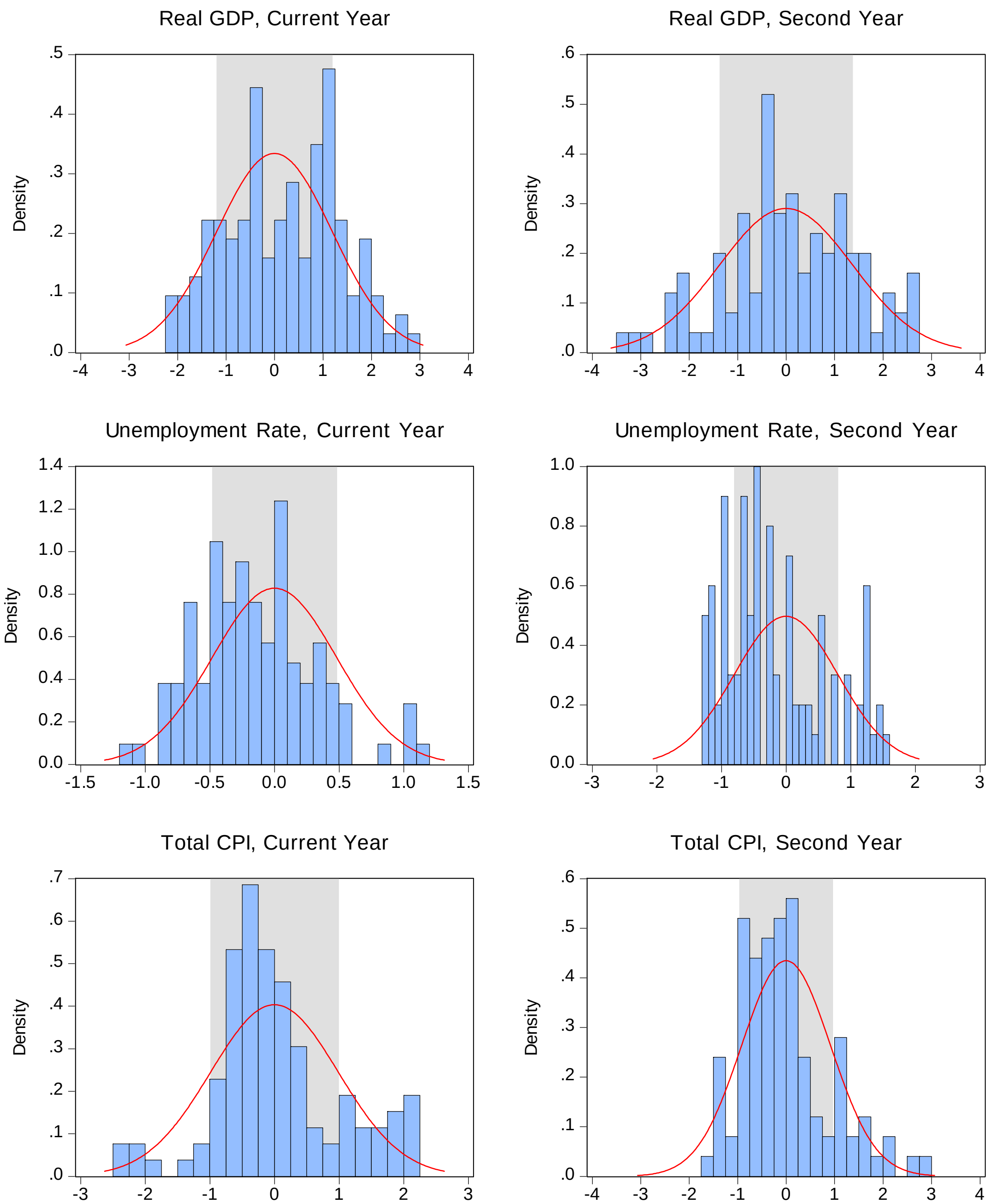


Figure 3
Summer Projection Errors -- Actual Distribution Vs. Normal Density
(shaded area equals zero plus or minus average root mean squared error)

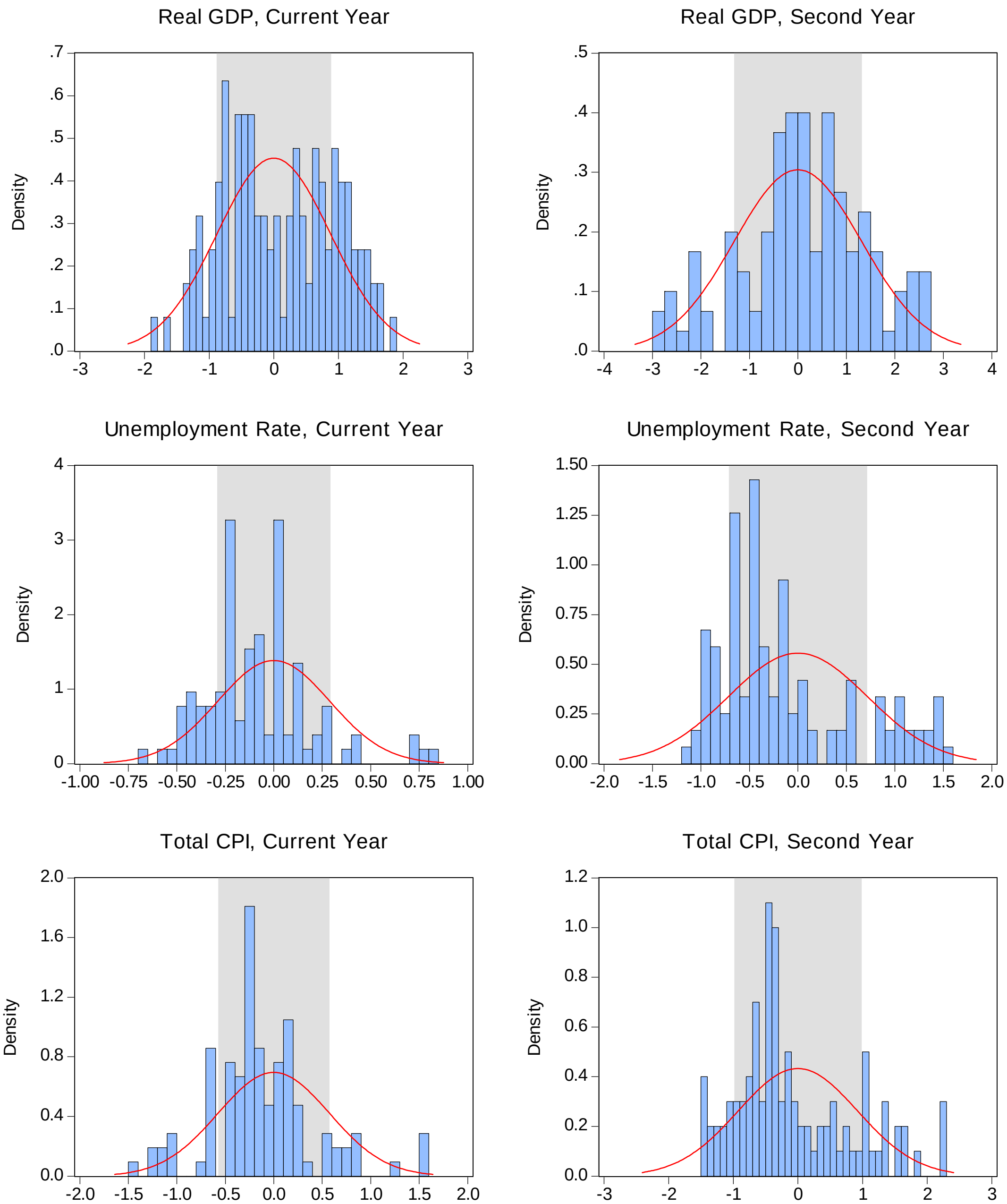
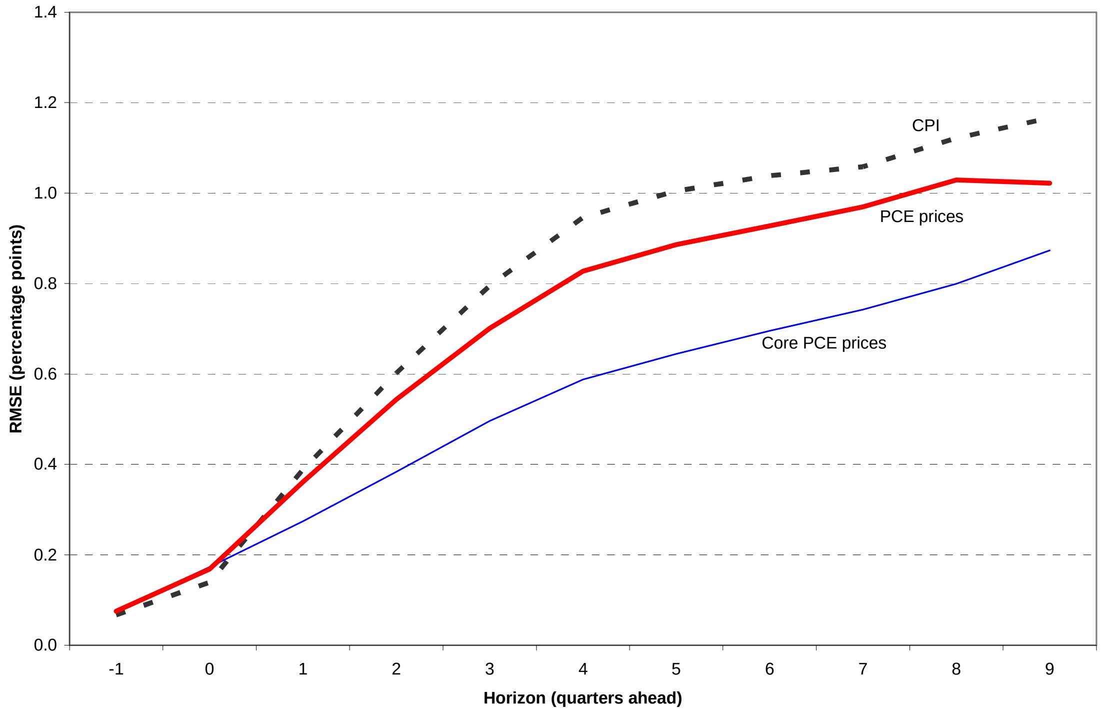


Figure 4:
Greenbook Inflation RMSEs, 1986-2006
(All 4-quarter changes; different measures)



Notes:

a) The CPI forecast errors used to generate this figure differ from those used to generate the results shown in Table 4 (and elsewhere in the paper) in two ways. First, the errors in this figure are generated using all eight Greenbook forecasts produced each year, rather than just the January, May, June and October forecasts. Second, the errors shown here are based on all four-quarter forecasts (that is, the q1-over-q1, q2-over-q2, q3-over-q3, and q4-over-q4 forecasts), and not just the q4-over-q4 forecasts.

b) The forecasts of the staff of the Federal Reserve Board of Governors are referred to as “Greenbook” forecasts as shorthand. In fact, forecasts of total and core PCE prices have only been included in the Greenbook since 2000; earlier forecasts are taken from unpublished Federal Reserve archives.