How Well Do Central Bankers Understand Fiscal Policy?

By Dean Croushore and Simon van Norden*

We create a real-time database of the Board staff's Greenbook forecasts of key fiscal policy variables and evaluate their properties. We find (1) limited evidence of forecast bias in the Greenbook forecasts, and find that they contain useful information beyond that in the CBO's forecasts; (2) the sign of past forecast errors helps to predict the sign of future errors; (3) the forecasts capture most of the variability of expenditures; (4) since the start of the Great Moderation, forecasts failed to anticipate most of the variation in government revenues, surpluses and structural surpluses.

JEL: E62, H68

Keywords: fiscal policy, deficits, forecasting, FOMC, Greenbook

* Croushore: Robins School of Business, 1 Gateway Road, University of Richmond, VA 23173, dcrousho@richmond.edu. van Norden: HEC Montréal, 3000 Chemin de la Côte Sainte Catherine, Montréal, QC, Canada H3T 2A7, simon.van-norden@hec.ca. The authors would like to thank Sdok Sambo, Kuy Sokha, and their Digital Divide Data team in Cambodia for their work in entering the data, as well as Nicholas Knowlton-Winch of CIRANO and Wendy Chan of the Bank of Canada for their research assistance. The authors would also like to thank seminar participants at the Joint Statistical Meetings in Miami Beach and Montreal, at the Central Bank of Cyprus, the Bank of Canada, the CIRANO Real-Time Workshop, the American Economic Association, and the University of Richmond, as well as Gitanjali Kumar and Valerie Ramey for their helpful comments.

The Zero Lower Bound on interest rates and subsequent experimentation with Quantitative Easing have powerfully occupied the attention of macroeconomists and central banks in recent years (and for good reason). At the same time, however, the Global Financial Crisis and the subsequent European Debt Crisis have only served to highlight another, more persistent change in the monetary policy environment: the increased variability of fiscal variables. For example, the extreme swings of U.S. fiscal policy in recent decades, from large deficits in the 1980s, to large projected surpluses at the end of the 1990s, to still-larger deficits thereafter, are without precedent in peacetime.

The potential interaction of monetary and fiscal policy gives the Fed staff strong motivation to forecast fiscal variables well; significant time and effort is invested, and there is discussion of fiscal policy in every FOMC Greenbook. A particularly striking example of the interaction between fiscal policy and monetary policy came in the late 1990s with the arrival of substantial federal government surpluses. At the time, projected surpluses suggested a possible future shortage of government bonds in financial markets, leading Fed economists to consider how to conduct monetary policy in the absence of federal government debt. The recession of 2001 and subsequent tax cuts eliminated this problem, but it is clear that the Fed was quite concerned about the potential supply of its main asset. More generally, understanding monetary policy requires us to understand how central bankers have perceived and anticipated the fiscal shocks they faced.

While there has been considerable work on the accuracy of central bank forecasts (such as those by the Federal Reserve Board's staff in the Greenbook) we are not aware of any that have examined fiscal variables. Instead, some of the best work on fiscal forecasts in recent years has been done on Eurozone data, due in part to the availability of suitable data sets. As we explain below, work on U.S. data has used forecasts that are perceived to have important defects. This paper begins to remedy that situation by documenting and analyzing a new coherent database of Federal Reserve Board forecasts of U.S. federal fiscal policy

variables. In doing so, it allows us to understand the extent to which monetary policy makers have been able to understand and anticipate fiscal changes, as well as how they learn over time about the trajectory of the federal government's fiscal balance.

The evaluation of fiscal forecasts and fiscal policy also raises a number of measurement-related issues. Evaluations are commonly based on currently available macroeconomic data. However, those data may differ in several ways from the information that was available to policymakers at the time. As Cimadomo (2011) notes, fiscal data are frequently revised. Others, such as Croushore (2011), note that GDP data are also frequently revised and business cycle turning points are identified only with a lag, making real-time considerations important. We therefore carefully match fiscal forecasts with contemporaneous data vintages of other key variables to allow us to properly understand the information available to policymakers. We believe this is the first paper to do so for U.S. fiscal forecasts. We also examine estimates and forecasts of the cyclically adjusted deficit to understand better how fiscal policy relates to perceptions of economic conditions.

We begin the paper in section I with a discussion of the literature on forecasts of fiscal policy, followed by section II, which describes the Greenbook data set and the data transformations we use. We formally evaluate the quality of the Greenbook forecasts in section III, testing them for bias, bias around elections, and inefficiency. We also compare the properties of the forecast errors of the Greenbook forecasts to those of the CBO. Section IV looks at variance decompositions of the forecast errors to measure the informativeness of the Greenbook forecasts, while Section V examines the distribution of forecast errors and how they compare across variables. We summarize the results and draw conclusions in section VI.

I. Literature on Fiscal Policy Forecasting

The literature on forecasting fiscal policy variables is sparse compared with that on forecasting monetary policy variables. Perhaps due to the relative importance of fiscal policy discipline in the Eurozone, much of the recent literature has examined fiscal policy forecasts in the European Union (EU), where the institutional framework has been quite different from that in the United States. We will therefore review fiscal forecasting separately for the U.S. and the EU to set the stage for our later analysis.

A. The U.S. Experience

Two official government agencies forecast U.S. government spending, revenues, and deficits—the CBO and the Office of Management and Budget (OMB). The CBO is a nonpartisan arm of the U.S. Congress, which is responsible for providing nonpolitical analysis of government budget issues. The OMB is part of the U.S. Treasury Department and works for the President to analyze his budget proposals. Researchers have compiled data sets to analyze both forecasts on an ad hoc basis, but there is no continuing program to update such data sets or to make them available to other researchers.

In their recent analysis of the CBO forecasts, Kliesen and Thornton (2012) show that the CBO's one-year-ahead forecasts are not significantly better than a random walk model (which assumes that next year's deficit will equal last year's deficit). The CBO's five-year projections are worse (though not statistically significantly worse) than the random walk model. It might not be a surprise that the CBO forecasts are worse in recessions than in expansions, as is likely true for all forecasters.

In their more comprehensive analysis of CBO, OMB, and Global Insight forecasts, Croushore and Hunt (2008) examine forecasts of deficits, revenues, outlays, and macroeconomic variables. They find that the forecasts are inefficient and

sometimes biased. The errors in the fiscal forecasts are attributable to poor forecasts of macroeconomic variables, including GDP, inflation, and unemployment. The results suggest that the government agencies would be better served by using private-sector forecasts of macroeconomic variables, rather than their own forecasts of those variables.

Other studies that examine both the CBO and OMB forecasts include Auerbach (1994), Auerbach (1999), and Plesko (1988). Auerbach (1994) shows that both CBO and OMB forecasts have generally been overly optimistic. Auerbach (1999) examines the revisions to the fiscal forecasts, finding that forecast revisions are serially correlated, suggesting inefficiency, especially for OMB forecasts. Plesko finds that long-horizon revenue forecasts are biased upwards, but most other forecasts are unbiased.

A few other studies have looked at particular aspects of fiscal forecasts. One study, Belongia (1988), compares the CBO's forecasts of deficits with those of the Council of Economic Advisers (CEA) and private-sector forecasts and finds no evidence of bias in the forecasts, though private-sector forecasts were more efficient than the CBO or CEA forecasts. Reischauer (1990), showed that the Gramm-Rudman-Hollings Act changed the nature of the OMB's summer forecasts, which were used to determine sequestration under the law, making them more optimistic (forecasting smaller deficits) than the OMB's winter forecasts, which did not affect sequestration. In contradiction to Plesko's results, Blackley and DeBoer (1993) find that forecasts of outlays were biased during Republican administrations, perhaps because those administrations used the forecasts as a bargaining tool. Campbell and Ghysels (1995) confirm Blackley and DeBoer's findings that the OMB's outlay forecasts are inefficient.

Compelling rationales for the bias and inefficiency of the CBO and OMB forecasts exist. The OMB is part of the government administration, and its forecasts are often used as a tactical weapon in political budget battles. The CBO is nonpartisan but is constrained to forecast revenues and expenditures according to

the *current law*, so it cannot condition on expected legislative changes. These inherent limitations create a void for researchers attempting to model or measure expected fiscal policy.

The Greenbook forecasts that we examine below are not unconditional forecasts: they are conditional on monetary policy assumptions. Improbable monetary policy assumptions will make fiscal policy forecasts unrealistic to the extent that those monetary assumptions affect forecast economic activity and the financing costs of the government debt. Given that previous studies have found Greenbook forecasts for economic activity to be quite good as *unconditional* forecasts, we expect such effects to be small. Thus, we expect the Greenbook forecasts to be of great interest. To our knowledge, the only previous study to have used Greenbook forecasts of fiscal variables is Auerbach and Gorodnichenko (2012), who used them only to construct measures of fiscal innovations and provide no direct analysis of their properties.[1]

B. Lessons from the European Union

Because of the Maastricht Treaty, researchers have devoted considerable effort to fiscal forecasts, beginning in the late 1990s. The fiscal forecasting literature, summarized by Leal et al. (2008), shows that some of the same issues of bias and inefficiency exist in Europe as they do in the United States. However, the EC's oversight of the forecasting process helps to control forecast errors. As Leal et al. note, "Most studies on forecast track records tend to signal that projections by the EC for European countries are the most accurate within international organisations publishing fiscal forecasts, due to its being an independent authority."[2] In contrast, Beetsma, Giuliodori and Wierts (2009) find that fiscal adjustments systematically fall short of forecast adjustments and that this shortfall increases with

[1]There are several important differences between their work and ours. Most notably, they use only one-quarter ahead forecasts for the growth rates of overall government spending and some of its components. We examine forecasts at multiple horizons for the level of federal government expenditures, receipts and other variables.

[2]See Leal et al. (2008), p. 350.

the forecast horizon. They also present evidence suggesting that as adjustment shortfalls accumulate, governments increasingly resort to creative accounting to mask the problem. Frankel (2011) finds that official forecasts of budget surpluses and overall growth are more (optimistically) biased in the case of Eurozone governments than for other nations he examines.

However, as is the case with the U.S. CBO, the EC is constrained to forecast based on "present policies," so its forecasts are not truly unconditional. Still, Artis and Marcellino (2001) find that there are not statistically significant differences in deficit/GDP forecasts for European countries between the IMF, the OECD, and the EC, where the former two presumably produce unconditional forecasts.[3]

II. Greenbook Forecasts—A New Data Set

To assess the Fed's ability to forecast fiscal variables, we first compiled fiscal forecasts from all Greenbooks from July 1966 to December 2006.[4] The Greenbook reports the Federal Reserve Board staff's forecasts before every FOMC meeting (which take place at least twice per quarter).[5] We examine the first and last Greenbook of each quarter to obtain a consistent data set with eight forecasts of quarterly data per year.

In each Greenbook, we gathered all the quarterly federal fiscal forecasts and reports of past data that are available for receipts, expenditures, the surplus, the high-employment budget surplus (HEB), a version of HEB based on a 6.1 percent or 6.0 percent natural rate of unemployment (which we call HEB6), the current

[3]To some extent, of course, the findings of bias and inefficiency of forecasts may depend on assumptions about the symmetry of the loss function. For example, Elliott, Komunjer and Timmermann (2005) find that IMF and OECD forecasts of G7 budget deficits are not rational under the assumption of symmetric loss but may be rationalized under asymmetric loss.

[4]The underlying data are available at the websites of the Federal Reserve Bank of Philadelphia and the Federal Reserve Board of Governors. See the Appendix to Croushore and van Norden (2014) for details. As with other FOMC briefing materials, Greenbooks are not released for at least five years. We end our sample before the start of the Global Financial Crisis; complete data covering the fiscal response to the crisis will not be available for some years yet.

[5]Do not confuse the Board staff's Greenbook with that of the U.S. Treasury; the Treasury's "Greenbook" is an annual publication containing the tax proposals in the President's budget proposal. That is not the Greenbook we use; ours is the Federal Reserve Board staff's forecast for the economy, which mainly focuses on overall macroeconomic variables, but also includes forecasts of federal fiscal variables, including the deficit or surplus, receipts, expenditures, and the high-employment budget balance.

and capital account surplus (which was introduced in 1996), the unemployment rate, nominal output, and real output.[6] The HEB variables are designed to measure the cyclically-adjusted budget surplus.

The occasional redefinition of some of our data series caused some complications. For example, beginning in 1996, overall government spending was replaced by government consumption expenditures and investment. Government spending on investment was removed from expenditures, but depreciation of capital was added. So, in periods when government investment exceeded depreciation, government expenditures were revised downwards. This caused both the surplus as well as GDP to be revised upwards. Another important change came in October 1999, when the BEA began treating government expenditures on software as investment. Again, this caused downward revisions to government expenditures and upward revisions to the surplus. Also, beginning in the early 1980s, HEB was based on a 6 percent natural rate of unemployment, but before that, the assumed natural rate of unemployment varied as it drifted upwards from an initial 4 percent rate.

Our primary data sources were page scans of the Greenbook independently published by the Federal Reserve Board and the Real-Time Data Research Center at the Federal Reserve Bank of Philadelphia.[7] After initial data entry and error-checking by a commercial firm, we compared some series (e.g., unemployment) against known values from other sources and checked the rest against the original PDF files. We believe our data to be at least as accurate as other published sources and our error rate to be less than 0.05%. The Appendix to Croushore and van Norden (2014) provides more details on the construction of our data set. Figure 1 shows a sample Greenbook page. Each variable in it can be represented as a string of estimates for past quarters (horizons -1, -2, etc.), the current quarter (horizon 0), and future quarters (horizons 1, 2, etc.).

[6]All the fiscal variables are reported on a National Income and Product Accounts (NIPA) basis, rather than a fiscal-year basis.

[7]See the Federal Reserve Board website for FOMC Historical Materials and the Philadelphia Fed's

The forecast horizons reported in the Greenbook varied considerably over time as shown in Figure 2. Greenbook forecasts generally go to the end of a calendar year; as the year progresses, we see somewhat fewer quarters of forecasts and somewhat more quarters of historical data. Both then change abruptly once a year when forecasts for the following calendar year are added. The earliest Greenbooks we recorded might contain only two quarters of forecasts and four quarters of current and historical estimates; none contained estimates more than 12 quarters ahead or into the past. As we examine longer forecast horizons (particularly those more than four quarters ahead), our sample is progressively drawn from more recent Greenbooks. For that reason, when comparing results across different forecast horizons, we sometimes restrict the sample period. For forecast horizons up to four, all of our series have at least one forecast per year from the first meeting in 1974Q4 onwards.[8]

After compiling the raw data, we normalized all fiscal variables, dividing them by the corresponding Greenbook values for nominal output (GNP before 1992, GDP from 1992 on).[9] This makes it easier to compare values across time. One such comparison is given by the string diagram in Figure 3, which shows the budget surplus as a share of GDP (or GNP).

String diagrams concisely show how forecasts evolve over time. For example, the early 1990s was a period when projections of steadily improving fiscal balances were met with a steadily deteriorating deficit. By the late 1990s, however, projections of roughly constant deficits and surpluses missed a sustained fiscal improvement. After 2001, however, we see a return to a pattern of persistently overoptimistic projected surpluses. This pattern looks different from the behaviour we see in the first half of the sample, something we investigate below.

Real-Time Data Research Center web site.

[8]Expenditures, receipts, HEB, and HEB6 typically have the shortest forecast horizons.

[9]Note that our series were recorded in levels, not growth rates.

III. Evaluating the Forecasts

Forecast evaluation requires a comparison of forecasts with a measure of outcomes. As the real-time literature shows (see Croushore (2011)), the revision of published macroeconomic data means that the choice of outcome measures (also called realized or actual values) may affect our results.

To evaluate the Greenbook forecasts, we considered five alternative measures of outcomes: (1) the last value published in the Greenbook (*last*); (2) the first officially published estimate (*initial*); (3) the officially reported value as of one year after the initial release (*one year*); (4) the last reported value before a benchmark revision of the National Income and Product Accounts NIPA (*prebenchmark*); and (5) the "current" official estimate (*current vintage*, which was current as of December 2012). The initial release, one-year release, and prebenchmark release of each variable come from the ALFRED database at the Federal Reserve Bank of St. Louis.

The importance of the differences between these alternative measures of outcomes varied considerably. For example, Figure 4 shows the results for government expenditures. Generally speaking, the redefinitions of the NIPA federal government accounts in 1999 had an economically large impact on receipts and expenditures, though not on the surplus. Other benchmark revisions were sometimes important, as were more regular revisions in some cases. On the other hand, unemployment rates underwent no substantial revisions. No statistical agency publishes estimates for our structural deficit measure, HEB; we therefore just compare its forecasts with the last reported value (*last*).

The combination of regular and benchmark revisions sometimes causes our current measures of fiscal variables to be very different from the earlier measures, particularly for both expenditures and receipts.[10] Benchmark revisions in particular may cause a researcher to find widespread evidence of forecast bias simply

[10]Revisions to the current vintage for expenditures and receipts have been largely offsetting, so revisions to the reported surplus have been relatively smaller.

because the precise definition of the series has changed since the forecasts were made, so that the currently published series give a distorted view of the forecast's performance. We take care to avoid such problems in the analysis which follows, most commonly by using prebenchmark estimates as our outcome measure.[11]

The Greenbook forecasts have a reputation for excellence in forecasting macroeconomic variables, as Romer and Romer (2000) show. Are they as good at forecasting fiscal policy variables? To find out, we ran some simple tests for bias and inefficiency.

A. Bias

To test for forecast bias, we examine forecasts covering horizons longer than the frequency of the observations, so the tests are subject to the standard overlapping observations problem.[12] We adjust for this by correcting the covariance matrix via Newey-West methods, using the lag length equal to the forecast horizon minus one. The results of the tests are summarized in Table 1. The table shows p-values for the null hypothesis of no bias for three different forecast horizons (zero, two, and four quarters ahead), four different concepts of realizations (last, initial, one year, and prebenchmark), two different meeting times during the quarter (first and last) and six different variables (surplus, expenditures, receipts, HEB, HEB6, and the unemployment rate.)[13]

There is no significant evidence of bias for forecasts of the budget surplus using any of the four outcome measures. Expenditure forecasts are significantly biased (forecasts exceeded realizations, on average) at a zero-quarter horizon, but not for longer horizons. The evidence for forecasts of receipts is mixed, with weaker

[11] This means omitting forecasts made just before a benchmark change for which official estimates were published only after the change.

[12] A basic test of forecast performance is the Mincer-Zarnowitz test, regressing the realized values of a variable on a constant and the forecasts. If the forecasts are unbiased, the constant term should be zero and the coefficient on the forecasts should equal 1. However, Mankiw and Shapiro (1986) show that in small samples (which is the case here), such tests may reject too often because the right-hand side variable is often autocorrelated and thus correlated with lags in the error term. Instead, a zero-mean forecast error test covers the same concept (and is a necessary condition for unbiasedness) without being subject to the small-sample bias.

[13] We ignore the current vintage realizations here because of the redefinition problem described above.

Table 1—Summary Results of Bias Tests

		Surplus		Expenditures		Receipts	
Horizon	Concept	First	Last	First	Last	First	Last
0	Last	0.56	0.08	0.02	0.03	0.04	< 0.01
	Initial	0.33	0.94	< 0.01	< 0.01	0.28	0.04
	One Year	0.60	0.71	< 0.01	< 0.01	< 0.01	< 0.01
	Prebenchmark	0.20	0.57	< 0.01	0.01	0.29	0.07
2	Last	0.37	0.40	0.64	0.55	0.18	0.21
	Initial	0.77	0.86	0.77	0.66	0.46	0.54
	One Year	0.63	0.70	0.24	0.17	0.04	0.06
	Prebenchmark	0.84	0.93	0.65	0.55	0.37	0.44
4	Last	0.23	0.22	0.84	0.78	0.03	0.04
	Initial	0.37	0.36	0.75	0.75	0.08	0.09
	One Year	0.31	0.31	0.60	0.51	< 0.01	< 0.01
	Prebenchmark	0.42	0.42	0.89	0.82	0.09	0.10

		HEB		HEB6		Unemployment	
Horizon	Concept	First	Last	First	Last	First	Last
0	Last	< 0.01	< 0.01	0.49	0.39	< 0.01	0.05
	Initial					< 0.01	0.40
	One Year					< 0.01	0.40
2	Last	< 0.01	0.02	0.30	0.50	0.06	0.03
	Initial					0.09	0.06
	One Year					0.09	0.06
4	Last	< 0.01	< 0.01	0.12	0.12	0.13	0.11
	Initial					0.16	0.15
	One Year					0.16	0.15

Note: The figures shown are p-values for tests of the null hypothesis that the mean forecast error is zero. Calculations use heteroscedasticity and autocorrelation robust standard errors with the number of lags equal to the forecast horizon minus one. The sample period is 1974:Q4 to 2006:Q4, except for HEB6, for which the sample begins in 1981:Q1. *First* and *Last* refer to the first and last FOMC meetings of each quarter.

evidence of bias at the shortest horizon but stronger evidence at longer horizons (with forecasts exceeding realizations, on average). HEB forecasts are biased for all horizons (again with forecasts exceeding realizations, on average) while there is never significant evidence of bias for HEB6, suggesting that the "drift" in the benchmark rate of unemployment prior to the early 1980s is responsible for the bias. The unemployment rate for the current-quarter shows bias only for the first meeting of the quarter.[14] At longer horizons, evidence of bias is marginal.

To understand why the receipt forecasts might be biased, we plot the four-quarter-ahead forecast against the one-year realized value in Figure 5. It shows that there is some tendency for the forecasts to exceed the realized value one year later. Such a tendency is not apparent in either the surplus forecasts or the expenditure forecasts, however. A time-series plot (not shown) makes it clear that the forecast errors in receipts were particularly large in the late 1990s and early 2000s, when the Greenbook persistently forecasted a rise in receipts that did not materialize. In this period, the Greenbook (and other forecasters) did not foresee the tax cuts that would be put in place, as well as the slowdown in the tech sector and the economy in 2000 and 2001.

Some researchers criticize tests of the mean forecast error for their sensitivity to large outliers and lack of power in some situations. We therefore also performed tests of the null hypothesis that the median forecast error was zero, following Campbell and Dufour (1991) and Campbell and Ghysels (1995).[15] Table 2 shows the p-values of the sign test statistic of the null hypothesis that forecast errors have a median of zero, while Table 3 shows the p-values of the Wilcoxon signed-rank test statistic of the null hypothesis that forecast errors are symmetrically distributed around zero.

The results, while similar in many ways to those previously discussed, provide

[14]Recall that this is a "nowcast" of a *quarterly average* unemployment rate. By the last FOMC meeting of the quarter, unemployment figures will have already been published for one or two of the three months in the quarter.

[15]These tests control for serial correlation in forecast errors caused by overlapping forecasts and allow for exact inference in small samples.

Table 2—Sign Tests of Forecast Errors for Various Outcome Measures

Sign Test		**SURPLUS**		**EXPEND**		**RECEIPTS**	
Horizon	Outcome	First	Last	First	Last	First	Last
0	Last	0.473	0.512	0.589**	0.566	0.566	0.605**
	Initial	0.481	0.504	0.667***	0.620***	0.496	0.566
	1 Year	0.473	0.504	0.651***	0.651***	0.581*	0.628***
	PreBenchmark			0.643***	0.612***	0.543	0.620***
2	Last	0.381	0.429	0.571	0.556	0.444	0.429
	Initial	0.381	0.381	0.444	0.571	0.413	0.429
	1 Year	0.333**	0.413	0.587	0.556	0.476	0.54
	PreBenchmark	0.349**	0.333**	0.556	0.54	0.429	0.397
4	Last	0.29	0.355	0.194***	0.194***	0.29	0.323
	Initial	0.29	0.323	0.194***	0.161***	0.258**	0.29
	1 Year	0.29	0.258**	0.194***	0.194***	0.355	0.645
	PreBenchmark	0.323	0.29	0.194***	0.161***	0.29	0.323
Sign Test		**HEB**		**HEB6**		**UNEMP**	
Horizon	Outcome	First	Last	First	Last	First	Last
0	Last	0.581*	0.628***	0.426	0.442	0.767***	0.791***
	Initial					0.667***	0.55
	1 Year					0.667***	0.558
2	Last	0.619*	0.651**	0.333**	0.397	0.730***	0.746***
	Initial					0.730***	0.714***
	1 Year					0.730***	0.714***
4	Last	0.548	0.581	0.355	0.355	0.742**	0.806***
	Initial					0.710**	0.742**
	1 Year					0.710**	0.742**

Note: The figures shown are the proportion of forecast errors > 0. Asterisks indicate the p-values associated with tests of the null hypothesis that the median forecast error is zero (*/**/*** indicate p-values less than 10/5/1 %). Test size is corrected for overlapping forecast horizons: see Campbell and Ghysels (1995) for details. The sample period is 1974Q4 to 2006Q4, except for HEB6, for which the sample begins in 1981Q1.

TABLE 3—WILCOXON SIGNED-RANK TESTS OF FORECAST ERRORS FOR VARIOUS OUTCOME MEASURES

Signed-Rank Test		**SURPLUS**		**EXPEND**		**RECEIPTS**	
Horizon	Outcome	First	Last	First	Last	First	Last
0	Last	4357	4805	5315*	5080*	5129*	5288*
	Initial	3785	4081	5562*	5332*	4769	4962*
	1 Year	3971	4276	5816*	5783*	5240*	5494*
	PreBenchmark	3534	3677	5525*	5283*	4776	5054*
2	Last	725	714*	783	804	875	867
	Initial	638*	614*	778	808	794	787
	1 Year	640*	651*	875	1124	928	915
	PreBenchmark	631*	608*	826	851	823	785
4	Last	133*	142	43*	44*	80*	90*
	Initial	126*	132*	47*	43*	80*	89*
	1 Year	129*	132*	47*	51*	88*	103*
	PreBenchmark	124*	119*	42*	45*	80*	90*

Signed-Rank Test		**HEB**		**HEB6**		**UNEMP**	
Horizon	Outcome	First	Last	First	Last	First	Last
0	Last	5365*	5666*	2967*	3058*	6045*	5747.5***
	Initial					5743.5***	4709
	1 Year					5739*	4704
2	Last	1175	1169	567*	503*	1400*	1448.5***
	Initial					1388.5**	1419*
	1 Year					1387.5**	1417*
4	Last	189	218	95*	97*	356	386.5**
	Initial					360.5	383.5**
	1 Year					360.5	383.5**

Note: The figures shown are the Wilcoxon Signed-Rank statistics associated with tests of the null hypothesis that the median forecast error is zero and its distribution is symmetric. Asterisks indicate the p-values associated with the tests (*/**/*** indicate p-values less than 10/5/1 %). Test size is corrected for overlapping forecast horizons: see Campbell and Ghysels (1995) for details. The sample period is 1974Q4 to 2006Q4, except for HEB6, for which the sample begins in 1981Q1.

more evidence of unconditional forecast bias. Nowcasts for expenditures and receipts continue to show strong evidence of bias, which is now also found for many of the four-quarter forecasts as well. Perhaps surprisingly, those horizons with the strongest evidence of bias in forecasts of expenditures or receipts are precisely those where evidence of bias in the forecast surplus is the weakest, suggesting the possibility of offsetting biases. The reverse is also found; at a two-quarter horizon, there is little sign of bias for expenditures or reciepts, but quite a bit of evidence of bias for the surplus. There is somewhat less evidence of bias than before for HEB, but somewhat more for HEB6. We also find significant evidence of bias in the unemployment rate forecasts at all horizons. Data revisions appear to play only a minor role in accounting for revisions, with the use of *Last* estimates typically providing as much evidence of bias as the *Initial* estimates; this is consistent with the hypothesis that data revisions for these series are themselves unforecastable.

As mentioned above, all of our fiscal forecasts are expressed as ratios relative to forecast values of nominal output (GNP or GDP.) This implies that our calculated forecast errors are influenced by the forecast error of both the fiscal variables and those of nominal output. We investigated the importance of the latter by instead scaling the fiscal forecasts by the realized values of nominal output. This had no detectable impact on the results for forecasts of the surplus or HEB. (The correlations between these two measures of forecast errors exceeded 0.99 for every forecast horizon.) However, the alternative scaling modified results for the revenue and expenditure forecasts somewhat. Particularly at longer horizons, this tended to reduce the forecast values of both series, thereby lowering their mean forecast errors by about 0.002 (i.e. two-tenths of one percent of output) at a four-quarter forecast horizon. In the case of revenues, this effectively eliminated the significant evidence of forecast bias. However, it had the opposite effect on expenditure, producing significant evidence of a negative forecast bias at longer horizons.[16]

[16]It should be noted that these results do not contradict the existing literature which finds no evidence of bias in Greenbook forecasts of output growth. That literature focuses on real output, not nominal,

While our forecast bias results for the fiscal variables are new, some authors have previously tested for bias in the Greenbook unemployment rate forecasts. Their results have been mixed: Baghestani (2008) uses forecasts made from 1983 to 2000 and finds significant evidence of bias (the Board staff tended to overpredict the unemployment rate) that increases with the forecast horizon, while Clements, Joutz and Stekler (2007) use data from 1974–2000 and find no significant evidence of bias. These conflicting results could indicate that the bias has been greater in more recent years. This led us to examine whether our results were consistent over time. We do so by splitting the sample in half, with one sample from 1974Q4 to 1990Q4 and the other from 1991Q1 to 2006Q4. Results in Tables 4 and 5 show the results, comparing full-sample results from Tables 2 and 3 to the 1974-1990 and 1991-2006 samples.

The results show that evidence of forecast bias for the fiscal variables in the 1974-1990 period largely mirrors that in the full sample, but that there is no evidence of forecast bias in the 1991-2006 period at any forecast horizon for any series other than the unemployment rate. In the 1974-1990 period, HEB forecasts were most frequently too high (i.e. forecast structural surpluses were too optimistic) while those of HEB6 were more frequently too low. The former is consistent with the *ex post* upward revisions to the benchmark rate of unemployment used to construct HEB during this period. The latter, which reflects a shorter sample period, implies that structural surplus estimates *conditional on a 6% rate of unemployment* tended to be higher in retrospect than had been forecast. Forecast errors for the surplus were most commonly negative (outcomes were typically higher than expected) with the frequency seeming to increase with the forecast horizon. This comes despite the absence of any evidence of mean bias in Table 1 above. There is also much less evidence of bias in the forecast errors for receipts or expenditures, with both tending towards more positive forecast errors at the

and examines growth rates, not levels. Furthermore, it is the inverse of nominal output that enters into our calculations. Jensens Inequality implies that if the forecast of the level of a variable is unbiased, the forecast of its inverse will generally be biased.

TABLE 4—NONPARAMETRIC TESTS OF FORECAST BIAS: SURPLUS, EXPENDITURES, AND RECEIPTS

	Full Sample		1974Q4-1990Q4		1991Q1-2006Q4	
SURPLUS						
Horizon	> 0	Signed-Rank	> 0	Signed-Rank	> 0	Signed-Rank
0L	0.442	3677	0.385*	810*	0.500	1047
0F	0.411*	3534	0.369**	761**	0.453	1035
1L	0.391**	3049**	0.359**	724**	0.429	810
1F	0.398**	3014***	0.406	718**	0.397	795
2L	0.333**	608**	0.290*	113**	0.323	170
2F	0.349**	631**	0.290*	114**	0.355	188
3L	0.357	236**	0.300	41**	0.350	67
3F	0.310*	258**	0.200**	42*	0.350	63
4L	0.290	119**	0.267	21	0.200	30
4F	0.323	124*	0.267	23	0.200	30
EXPENDITURES						
Horizon	> 0	Signed-Rank	> 0	Signed-Rank	> 0	Signed-Rank
0L	0.612***	5283**	0.692***	1598***	0.531	1062
0F	0.643***	5525***	0.692***	1559***	0.594	1184
1L	0.555	4576	0.563	1279	0.540	972
1F	0.570*	4604	0.547	1222	0.587	1031
2L	0.540	851	0.581	294	0.484	191
2F	0.556	826	0.581	280	0.516	182
3L	0.452	331	0.550	101	0.400	67
3F	0.429	326	0.500	89	0.400	64
4L	0.161***	45***	0.000***	0***	0.333	31
4F	0.194***	42***	0.067***	1***	0.333	34
RECEIPTS						
Horizon	> 0	Signed-Rank	> 0	Signed-Rank	> 0	Signed-Rank
0L	0.620***	5054**	0.631**	1443**	0.609*	1103
0F	0.543	4776	0.554	1254	0.531	1161
1L	0.508	3842	0.563	1045	0.444	844
1F	0.508	3712	0.578	1010	0.429	808
2L	0.397	785	0.452	190	0.323	165
2F	0.429	823	0.516	193	0.387	187
3L	0.333	303	0.350	65	0.25	64
3F	0.429	294	0.550	56	0.3	67
4L	0.323	90***	0.133**	4***	0.267	38
4F	0.290	80***	0.133**	3***	0.333	39

Note: This table presents nonparametric tests for forecast bias. Forecast errors were calculated using our preferred outcome measures for each series, $PreBenchmark$ for SURPLUS, EXPEND and RECEIPTS. Forecast horizons increase moving down the table, with $Horizon$ giving the forecast horizon in quarters with the suffix L for the Last meeting of a quarter and F for the First. The $FullSample$ is 1974Q4 to 2006Q4. The figures shown in columns headed '> 0' are the fraction of forecast errors that are greater than or equal to zero. Those shown in columns headed 'Signed-Rank' are the Wilcoxon Signed-Rank statistics for the null hypothesis that the forecast errors are symmetrically distributed with a median of zero. Asterisks indicate the p-values for tests of the null hypothesis that the median forecast error is zero, with */**/*** corresponding to p-values less than 10/5/1%. Test size is corrected for overlapping forecast horizons: see Campbell and Ghysels (1995) for details.

TABLE 5—NONPARAMETRIC TESTS OF FORECAST BIAS: HEB, HEB6, AND UNEMPLOYMENT RATE

	Full Sample		1974Q4-1990Q4		1991Q1-2006Q4	
			HEB			
Horizon	> 0	Signed-Rank	> 0	Signed-Rank	> 0	Signed-Rank
0L	0.628***	5666***	0.754***	1790***	0.500	1049
0F	0.581*	5365***	0.692***	1686***	0.469	1023
1L	0.563	4692	0.703***	1579***	0.413	782
1F	0.586**	4842*	0.750***	1616***	0.429	826
2L	0.651**	1169	0.774***	378**	0.419	176
2F	0.619*	1175	0.774***	357*	0.387	183
3L	0.595	418	0.750**	141	0.350	70
3F	0.643	498	0.800***	143	0.400	76
4L	0.581	218	0.733	77	0.333	33
4F	0.548	189	0.733	77	0.333	35
			HEB6			
Horizon	> 0	Signed-Rank	> 0	Signed-Rank	> 0	Signed-Rank
0L	0.442	3058***	0.385*	545***	0.500	1049
0F	0.426	2967***	0.385*	538***	0.469	1023
1L	0.391**	2504***	0.359**	508***	0.413	782
1F	0.438	2677***	0.453	572***	0.429	826
2L	0.333**	503***	0.258**	75***	0.419	176
2F	0.397	567***	0.419	118**	0.387	183
3L	0.381	229**	0.350	40**	0.350	70
3F	0.381	248**	0.350	38**	0.400	76
4L	0.355	97**	0.200	6***	0.333	33
4F	0.355	95**	0.200	6***	0.333	35
			UNEMP			
Horizon	> 0	Signed-Rank	> 0	Signed-Rank	> 0	Signed-Rank
0L	0.558	4704	0.585	1308	0.531	1048
0F	0.667***	5739***	0.662***	1419**	0.672***	1468***
1L	0.641***	5665***	0.609*	1373**	0.683***	1488***
1F	0.703***	5699***	0.641**	1354**	0.778***	1531***
2L	0.714***	1417**	0.742***	357*	0.742***	373**
2F	0.730***	1387**	0.710**	342	0.774***	366**
3L	0.762***	665**	0.750**	156	0.800***	165*
3F	0.738***	637*	0.750**	164*	0.750**	164*
4L	0.742**	383**	0.800**	101*	0.800**	93
4F	0.710**	360	0.733	88	0.800**	93

Note: This table presents nonparametric tests for forecast bias. Forecast errors were calculated using our preferred outcome measures for each series (*Last* for HEB and HEB6, and $CurrentVintage$ for UNEMP. Forecast horizons increase moving down the table, with $Horizon$ giving the forecast horizon in quarters with the suffix L for the Last meeting of a quarter and F for the First. The $FullSample$ is 1974Q4 to 2006Q4 (except for HEB6, which begins in 1981Q1). The figures shown in columns headed '> 0' are the fraction of forecast errors that are greater than or equal to zero. Those shown in columns headed 'Signed-Rank' are the Wilcoxon Signed-Rank statistics for the null hypothesis that the forecast errors are symmetrically distributed with a median of zero. Asterrisks indicate the p-values for tests of the null hypothesis that the median forecast error is zero, with */**/*** corresponding to p-values less than 10/5/1%. Test size is corrected for overlapping forecast horizons: see Campbell and Ghysels (1995) for details.

shortest horizons and negative ones at longer horizons.

Forecasts for the unemployment rate were unusual in that they showed strong evidence of bias in both portions of the sample. We find that forecast errors for the unemployment rate are zero or greater more than 70% of the time at most forecast horizons, implying that the Greenbook forecasts are typically too pessimistic. This therefore reinforces the results reported by Baghestani (2008) and contrasts with those of Clements, Joutz and Stekler (2007). These results are also not inconsistent with those shown in Table 1, although that evidence of bias weakened as forecast horizons increased. Figure 6 compares the forecast unemployment rates (from the first FOMC meeting of each quarter) with outcomes. They show a pattern consistent with forecasts that explain very little of the observed variation in outcomes; forecasts steadily underestimated actual unemployment during downturns and overestimated it during recoveries. The fact that recoveries last longer than downturns may in turn explain why forecast errors were so frequently positive throughout the sample.

The results suggest that Greenbook forecasts of the fiscal variables show significant forecast biases, especially for expenditures, receipts, and unemployment at short horizons. On the other hand, it is likely that the Fed's staff spends much more time and attention on macroeconomic forecasts at longer horizons that may be more relevant to monetary-policy decision-making than on the fiscal "nowcasts." The evidence of bias also appears to be confined to the first half of the sample, with no evidence of bias for the fiscal variables after 1990.

B. Bias and Election Cycles

There has also been considerable interest in the potential for moral hazard to create forecast bias, particularly around elections. While there is some evidence of systematically optimistic forecasts in advance of elections, we might expect the Greenbook forecasts to be an exception as they are not publically released for at least five years, thereby reducing the direct moral hazard, and the Board is

typically portrayed as nonpartisan. We therefore also test for systematic forecast bias related to the U.S. presidential election cycle by regressing forecast errors on a constant and three dummy variables. These dummy variables are equal to one only in presidential election years ($ELECTION$), the year before presidential election years ($PRE-ELECTION$), and the year after presidential election years ($POST-ELECTION$).[17] For simplicity, we test only forecast errors using our "best" measure of forecast outcomes; prebenchmark estimates for expenditures and receipts, our current vintage for the unemployment rate, and the last Greenbook value for HEB, HEB6, and the overall surplus/deficit. To allow for sufficient degrees of freedom, we consider only forecast horizons from zero to four quarters ahead and test the period 1974Q4–2006Q4.

We do not report the results here for reasons of space, but they may be summarized as showing little or no evidence of forecast bias related to the election cycle. The joint hypothesis that all three dummy variables were equal to zero was rarely rejected at even the 10% significance level. What limited evidence of bias we could find was concentrated in nowcasts made in preelection years, where some series appeared to have a positive bias on the order of one-half of 1 percent of GDP. However, given the degree of "data snooping" involved in these tests, we found the evidence to be less than compelling.[18]

C. Inefficiency

Another important aspect of forecast performance is the efficiency of forecasts with respect to other variables that are in the information set of forecasters. In principle, a researcher could look for a relationship between forecast errors of any of the budget variables and data that were in the information set when each Greenbook forecast was produced. Because of the timing requirements, it is

[17] Standard errors for the estimated coefficients were corrected for serial correlation caused by overlapping forecast horizons using Hansen-Hodrick robust standard errors.

[18] We tested three dummy variables for each of seven series at 10 different forecast horizons for a total of 210 test statistics. The number of rejections of the null hypothesis that we found was roughly what we should have expected under the null hypothesis given the significance level of the test.

crucial that real-time data be used in such an exercise.

One finding in the literature is that forecasters sometimes do not adjust their forecasts properly for changes in monetary policy. Ball and Croushore (2003), for example, show that real output forecast errors from the SPF are correlated with past changes in monetary policy, as measured by the fed funds rate. (The advantage of using the fed funds rate in a test for inefficiency is that it is not revised.) We therefore examine our Greenbook forecast errors to see if they are inefficient with respect to changes in the fed funds rate. We use the four-quarter change in the fed funds rate ending in the quarter *before* the Greenbook forecast is made so that we are certain that the change in the fed funds rate was in the information set of the forecasters.

Table 6 shows the results of the efficiency tests. Note that we do not test for efficiency in instances in which we found non-zero-mean forecast errors in the test for unbiasedness earlier. (In such cases, the table cells simply read 'bias.') The results show no statistically significant evidence of inefficiency in the forecasts for any of the variables; the past change in monetary policy is not correlated with the forecast errors of these variables.[19] Thus, the Ball and Croushore (2003) results on the inefficiency of the SPF forecasts do not carry over to fiscal forecasts in the Greenbook.

There is also a substantial literature on another type of inefficiency in Greenbook forecasts. Starting with Scotese (1995), multiple studies have found that Greenbook forecast errors tend to be serially correlated.[20] We investigated this using the Sign and Signed-Rank tests for first-order serial correlation suggested by Campbell and Ghysels (1995). The results (not shown here to conserve space) strongly rejected the null hypothesis of forecast efficiency for all variables at the current-quarter and one-quarter horizons, although longer horizon forecasts

[19] Of course, other information that was available when the forecasts were made might be correlated with the forecast errors.

[20] Scotese (1995) proposes a rational model of such behavior in which forecasters attempt to reduce the variance of their forecasts as information arrives in order to appear more credible. "Anchoring," a well documented form of cognitive bias, would also produce such behavior.

showed less or no evidence of inefficiency.[21] The results were robust to the use of initial-release estimates to measure forecast errors (and so cannot be attributed to data revisions) and to splitting the sample into sub-periods from 1974-1990 and 1991-2006.[22]

TABLE 6—SUMMARY RESULTS OF EFFICIENCY TESTS

		Surplus		Expenditures		Receipts	
Horizon	Concept	First	Last	First	Last	First	Last
0	Last	0.21	0.14	bias	bias	bias	bias
	Initial	0.21	0.09	bias	bias	0.13	0.08
	One Year	0.60	0.56	bias	bias	bias	bias
	Prebenchmark	0.26	0.16	bias	bias	0.08	0.06
2	Last	0.88	0.99	0.19	0.22	0.19	0.35
	Initial	0.95	0.91	0.30	0.35	0.23	0.42
	One Year	0.88	0.75	0.34	0.38	bias	0.59
	Prebenchmark	0.98	0.84	0.17	0.19	0.08	0.17
4	Last	0.59	0.56	0.10	0.13	bias	bias
	Initial	0.66	0.63	0.09	0.10	0.09	0.11
	One Year	0.53	0.51	0.16	0.20	bias	bias
	Prebenchmark	0.54	0.52	0.09	0.12	0.10	0.15
		HEB		HEB6		Unemployment	
Horizon	Concept	First	Last	First	Last	First	Last
0	Last	bias	bias	0.45	0.92	bias	bias
	Initial					bias	0.20
	One Year					bias	0.20
2	Last	bias	bias	0.26	0.57	0.28	bias
	Initial					0.24	0.28
	One Year					0.24	0.28
4	Last	bias	bias	0.18	0.29	0.08	0.11
	Initial					0.08	0.11
	One Year					0.08	0.11

Note: The figures shown are p-values for tests of the null hypothesis that the coefficient on the lagged change in the federal funds rate is zero.

D. Forecast Comparisons

Another way to understand the efficiency of the Greenbook forecasts is to compare their performance with that of other forecasters. This kind of comparison is complicated by several factors, however. Many forecasters forecast the general

[21] The results indicated that sign of forecast errors tended to persist over time, with only one significant exception. In the case of HEB6, full-sample results as well as those for the early sample showed that the sign of forecast errors changed signs *more* frequently than predicted under the null hypothesis of forecast efficiency.

[22] The sole exception to this was UNEMP, where there was considerable evidence of inefficiency at horizons of four quarters and beyond, particularly in the 1991-2006 sample.

government sector rather than the Federal government. Some forecasters forecast variables on a budget-accounting basis rather than a National Income and Product Accounts basis. Many forecasters forecast only annual rather than quarterly totals, and their forecasts are updated less frequently than the Greenbook. Finally, other forecasts cover a much shorter historical period.

In light of these limitations, perhaps the best available comparison for the Greenbook forecasts are those produced by the CBO for the annual federal government surplus, expenditures and receipts. We take the first CBO forecast of each year and compare it to the corresponding Greenbook forecast by averaging the four quarterly Greenbook forecasts to compute the implied annual forecast.[23] Both sets of forecasts are compared in Table 7. Forecasts for the current and next calendar year were available from 1982 to 2006, except for expenditures and receipts where forecasts for the next calendar year were only available from 1990 onwards.

Table 7 compares the performance of the Greenbook and the CBO performance in a number of ways.[24] The first two lines simply report the root-mean-squared forecast errors. We see that CBO forecasts are slightly more accurate in two of the six cases. The third line tests the null hypothesis that the two forecasts have equal mean-squared forecast errors and reports the associated p-values.[25] We find that the Greenbook forecasts are significantly more accurate only for current-year forecasts of receipts and year-ahead forecasts of receipts and expenditures. There is no statistically significant difference in the accuracy of their forecasts of the surplus. Perhaps surprisingly given the apparently small difference in mean-squared forecast error, we also find that the CBO forecasts for year-ahead receipts

[23]CBO forecasts for fiscal variables were divided by their forecast values for nominal GNP or GDP to calculate the implied forecasts for output shares. Similarly, we averaged the Greenbook fiscal variables across the four quarters of each year before converting to output shares using the the Greenbook's output forecasts. The CBO forecasts were made in late January or early February of each year, except for 1996 when the forecast was made in May. Due to benchmark changes in the National Income and Product Accounts, we omitted those forecasts whose outcomes were affected by definitional changes. This had only a minor impact on our results.

[24]In interpretting these results, it should be recalled that these two forecast condition on distinctly different assumptions (as discussed above.)

[25]We use the modified Diebold-Mariano statistics proposed by Harvey, Leybourne and Newbold (1998).

are more accurate than the Greenbook's. The fourth line in the table shows how our conclusions change when we test the null hypothesis of equal expected absolute forecast error. Results are similar, although we now find that Greenbook performs significantly better for the year-ahead surplus but not expenditures.

TABLE 7—GREENBOOK VERSUS CBO

Variable	**SURPLUS**		**RECEIPTS**		**EXPEND**	
Horizon (Years)	**0**	**1**	**0**	**1**	**0**	**1**
RMSFE - Greenbook	0.00901	0.01396	0.00504	0.01026	0.00527	0.00918
RMSFE - CBO	0.00873	0.01658	0.00658	0.01008	0.00564	0.00962
H_0 : Equal Quadratic Loss	0.804	0.108	0.008	0.000	0.431	0.033
H_0 : Equal Absolute Loss	0.917	0.032	0.004	0.004	0.760	0.374
H_0 : GB encompasses CBO	0.075	0.225	0.923	0.001	0.200	0.528
H_0 : CBO encompasses GB	0.306	0.062	0.000	0.000	0.002	0.001

Note: $RMSFE$ indicates the Root-Mean-Squared Forecast Error. Figures shown for the null hypothesis of equal Quadratic or Absolute loss are p-values associated with the Diebold and Mariano (2002) test statistic of the corresponding null hypothesis. Figures in the final two rows test the null hypothesis of forecast encompassing using the statistic proposed by Harvey, Leybourne and Newbold (1998) and incorporate their proposed small-sample adjustment.

The final two lines of the Table provide the results of forecast encompassing tests. Forecast A is said to forecast encompass Forecast B if the *forecast errors* of A are uncorrelated with the *forecasts* of B. This implies that A is efficient in the sense that the information in B cannot be used to improve A. Our results show that we are able to strongly reject the null hypothesis that the CBO forecasts forecast encompass the Greenbook forecasts of receipts and expenditures (and we can reject the same hypothesis for the year-ahead surplus forecasts at the 10% level.) This implies that the Greenbook forecasts capture useful information that the CBO forecasts miss. One possible explanation for this is the CBO's requirement to forecast conditional on "current law," which forces them to omit information about expected legislative changes. However, we also find one instance (for year-ahead forecasts of receipts) in which the Greenbook forecasts clearly do not forecast encompass those of the CBO, indicating that CBO forecasters had valuable insights that the Board staff lacked.

IV. Forecast Uncertainty and Learning About the Present

Forecast bias and efficiency are interesting properties, but it is also useful for policymakers to understand how informative forecasts are likely to be. How successful are forecasts in capturing changes four quarters ahead? Two quarters? Zero? To measure this, we simply calculate the variance of the forecast errors as a share of the unconditional variance of the target series. Low values (close to zero) imply that forecasts are useful in the sense that they capture much of the movement in the series they attempt to predict. As values approach one, however, the forecasts capture less and less of the variation in the target variable.[26] Table 8 shows these ratios by forecast horizon, from the zero-quarter horizon for the last meeting of the quarter to the eight-quarter forecast for the first meeting of the quarter. As the target recedes into the future, we expect to see a steady rise in the relative variance of the forecast errors.

In all cases, nowcasts performed well, capturing the vast majority of the variation in the series. As forecast horizons lengthened, however, the deterioration in forecast performance varied widely, both across series and across the first and second halves of our sample. In the first half of the sample, forecasts for all series except HEB and HEB6 performed similarly, with forecast error variances consistently rising from less than 10% for the nowcasts to roughly 30% at a four-quarter horizon. HEB stands out as having a considerably higher relative forecast error at every horizon, reaching over 60% for the longest forecasts. Curiously, the forecast performance of HEB6 is roughly constant and independent of the forecast horizon.[27] Part of the difference is due to the changing benchmark unemployment rate used to calculate HEB through the 1970s, reflecting changing views of the natural rate of unemployment and of potential output.

In the second half of the sample, the results are quite different. Forecast errors

[26] Values greater than one imply a different kind of forecast inefficiency: one in which the user would be better off ignoring the forecast and simply using the unconditional mean of the target variable.

[27] HEB6 is not available prior to 1980, so the sample period used for it is substantially different from and shorter than that of the other series.

Table 8—Forecast Error Variance

Horizon	Expenditures	Receipts	Surplus	C&C Surplus	HEB	HEB6	Unemployment
1974Q4-1990Q4							
0L	0.047	0.075	0.056		0.179	0.141	0.008
0F	0.084	0.129	0.127		0.256	0.146	0.042
1L	0.105	0.189	0.181		0.314	0.152	0.067
1F	0.167	0.194	0.256		0.319	0.139	0.122
2L	0.149	0.330	0.284		0.413	0.117	0.155
2F	0.212	0.328	0.339		0.406	0.111	0.190
3L	0.197	0.173	0.196		0.383	0.143	0.229
3F	0.240	0.214	0.257		0.420	0.144	0.243
4L	0.221	0.207	0.206		0.448	0.093	0.315
4F	0.269	0.250	0.285		0.630	0.137	0.321
1991Q1-2006Q4							
0L	0.049	0.129	0.055	0.091	0.210		0.003
0F	0.059	0.141	0.074	0.114	0.243		0.011
1L	0.067	0.179	0.118	0.161	0.329		0.015
1F	0.069	0.228	0.132	0.159	0.314		0.030
2L	0.101	0.403	0.258	0.273	0.464		0.037
2F	0.118	0.421	0.291	0.288	0.451		0.054
3L	0.169	0.609	0.447	0.415	0.591		0.064
3F	0.195	0.630	0.491	0.450	0.586		0.098
4L	0.260	0.832	0.684	0.605	0.769		0.115
4F	0.304	0.844	0.750	0.667	0.786		0.154

Note: Forecast error variances are shown as a fraction of the unconditional variance of the underlying series over the period 1974Q4-2006Q4. Forecasts are taken from the first FOMC meeting in 1974Q4 until the last meeting in 2006Q4. Outcomes are measured as *last* for HEB, HEB6, and the Current and Capital Account Surplus, as *Prebenchmark* for Expenditures, Receipts, and the Surplus, and as *Current Values* for the unemployment rate. We omit the C&C Surplus in the first period as it is identical to the Surplus, and we omit HEB6 in the second period as it is identical to HEB.

for unemployment are the lowest of any series and are often less than half that of the values in the first half of the sample. All other series show a deterioration in forecast performance, with receipts and the surplus most seriously affected, particularly at the three-quarter and four-quarter horizons. This is particularly puzzling given that this was the period of the "Great Moderation" when the economy was relatively more stable.

An examination of the forecast errors shows that they were particularly large for the surplus in 1992 (about 2.0 percent of GDP) followed by large and sustained errors from 2001Q3 to the end of 2003 (always 2 percent or more of GDP). In both cases, deficits were substantially larger than expected. In large measure, this reflected a shortfall in receipts, which was then exacerbated in the latter period

by higher-than-expected expenditures. Both 1992 and the 2001-2003 period also featured similar, unusually large forecast errors in HEB. This suggests that these forecast errors were not primarily due to an unusually weak economy so much as they reflected a failure to anticipate government revenues and expenditures conditional on the state of the economy. That interpretation is also consistent with the relatively good performance in forecasting unemployment and the generally low volatility of the economy during the Great Moderation.

In summary, these results show that, while nowcasts for all variables were very informative, the usefulness of the forecasts varied considerably over time and across variables. In recent decades, forecasts of both actual and structural surpluses have been particularly difficult. These results also suggest that care should be taken in modeling the behavior of policymakers, as their expectations of fiscal policy may be quite different from what is subsequently observed. We return to this point below.

V. The Distribution of Forecast Errors

Another way to understand the relative performance of the various forecasts is to compare the distribution of forecast errors across forecast horizons and across variables. This is succinctly summarized in Figures 7 through 9. Each figure shows simplified box plots describing the distribution of forecast errors for each of 22 different forecast horizons. The plots (due to Tufte (1983)) summarize the information contained in a box plot with two vertical lines separated by a dot. The upper line runs from the 90th to the 75th percentiles while the lower line runs from the 25th to the 10th. The dot indicates the median. By overlaying box plots for two series, we see how their distributions compare and vary with the forecast horizon.[28] They provide information that is distinct from that shown

[28]Forecast errors are based on current vintage outcomes for the unemployment Rate, on Last Greenbook values for HEB, HEB6, and Anti-HEB, and on Pre-Benchmark values for all other series. Each box plot shown in these figures uses all the available observations for the series at the given forecast horizon. The number of observations therefore varies across forecast horizons and series. As a check, we also constructed figures based only on those observations for which forecast errors were available for all series

above in Table 8, which simply compared forecast error variances for the first and second half of the sample. They also go beyond the results in Table 4, which only tested the median of the distribution, by providing more information about the asymmetric risks of forecast errors.

Figure 7 compares the forecast error distribution for the surplus with that of expenditures (upper panel) and that of receipts (lower panel). While the former panel shows that the distribution of backcast errors for surpluses and expenditures are quite similar, the dispersion of forecast errors for the surplus rises with the forecast horizon more quickly than that of expenditures. Here we also see that the distribution is not symmetric; the odds of large positive errors (projected surpluses larger than actual outcomes) are greater than those of negative errors, particularly for forecasts more than two quarters ahead. This asymmetry does not appear to be accounted for by forecast errors in expenditures, whose distribution appears to be more symmetric, although some asymmetry is noticeable at some horizons. In the lower panel, we again see that the dispersion of forecast errors for receipts is much less than that of the surplus at all positive forecast horizons and the asymmetry of the distribution for expenditures is less pronounced than that for receipts.[29] This suggests that the risk of large surplus forecast errors, which is most commonly the result of overly optimistic surplus forecasts, is disproportionately due to overly optimistic revenue rather than expenditure forecasts.

Figure 8 compares the forecast error distribution for the structural surplus measure HEB6 with that of actual surplus (upper panel) and that of HEB (lower panel.) The latter clearly shows the effects of the upwards drift in the unemployment rate used to estimate the structural surplus. Forecast errors for HEB have a strongly skewed distribution. While the median error is never far from zero, positive errors exceeding 0.5% of GDP are roughly as common as negative errors

at all forecast horizons. This both greatly reduced the number of observations available for analysis and reduced the range of available forecast horizons (from -2 to 4 quarters instead of -4 to 6.) However, it confirmed that the results discussed below were not simply due to differences in sample periods.

[29]This asymmetry become even more pronounced when we limited our sample period to be the same for all three series and all forecast horizons. This also had the effect of reducing the median forecast errors for all three series and all forecast horizons

of 1.0–1.5%. This asymmetry is largely absent once the benchmark unemployment rate is held constant at 6% as we can see from the distribution for HEB6, which is roughly symmetric about its median. This is consistent with the experience from the mid-1960s to the early 1980s, where occasional sharp recessions produced large forecast errors in projected surpluses and led to upward revisions in the estimated "structural" rate of unemployment.

The upper panel shows that forecast errors for the actual surplus are more asymmetric than those for HEB6, which is what we might expect if the latter successfully abstract from an important source of asymmetric risks (i.e. the business cycle). It is also interesting to note that while the dispersion of forecast errors at the shortest horizons is greatest for the cyclically adjusted measure (HEB6), the opposite is true for forecast horizons of 2 quarters or more.[30] It seems that the business cycle's contribution to the surplus or deficit is a relatively more important source of uncertainty when nowcasting, but that this is swamped by uncertainty about the position in the cycle at horizons of more than a couple of quarters.

Figure 9 provides another perspective on the impact of business cycle uncertainty by comparing forecast errors for Unemployment with those of the cyclical budget surplus: Anti-HEB6.[31] Because they are measured in different units (the former in labor force shares, the latter in output shares) the scales of the two sets of distributions are not directly comparable. Despite this, the figure reveals some interesting features. First, although forecast errors for the unemployment rate have a significantly positive median (as noted in Table 5), there is little or no similar tendency in the implied forecast errors for the cyclical component of the surplus. On average, overly optimistic unemployment forecasts have tended to be offset by overly pessimistic *conditional* forecasts for receipts or expenditures.

[30] When the sample period is restricted to be the same for all forecast horizons and both series, the difference in dispersion of forecast errors across the two series at longer horizons increases, as the risk of large positive forecast errors in the surplus becomes more pronounced.

[31] Anti-HEB6 is simply the difference between the actual surplus and its estimated structural component HEB6.

Second, the forecast uncertainty for unemployment diminishes more abruptly as the forecast horizon shortens to zero than does that for the cyclical surplus. This is consistent with the minimal revision uncertainty and publication lags in official unemployment rate statistics, particularly when compared with the publication of the government sector of the National Accounts. Third, both variables show considerable asymmetry about their median forecast errors, with large negative errors more commonly than positive errors of a similar size. Recall that negative errors imply overly optimistic unemployment rate forecasts but overly pessimistic forecasts of the cyclical surplus.

VI. Summary and Conclusions

The goal of this paper was to better understand the Federal Reserve Board's ability to understand and anticipate changes in fiscal variables. To do so, we assembled a new data set containing a complete set of Greenbook fiscal forecasts spanning many decades and complete business cycles.

Our analysis highlighted both positive and negative aspects of the forecasts' performance. On the positive side, forecasts of both the surplus and the structural surplus (HEB6) appear to be unbiased and efficiently incorporate information from monetary policy varibles. Greenbook forecasts are in several cases slightly better than those of the Congressional Budget Office, both in terms of mean-squared errors and in terms of forecast-encompassing. Median forecast errors were never significantly different from zero for any of our fiscal variables in the latter half of our sample.

On the negative side, near-term forecasts of both government receipts and expenditures showed evidence of bias, and most variables had median forecast errors different from zero over the 1970s and 1980s. All short-term forecasts also seemed inefficient in the sense that the sign of past forecast errors helped to predict the sign of future errors. We also found that unemployment rate forecasts seemed particularly biased, with the Board staff forecasts consistently overpredicting the

rate.

More generally, we also found that the quality of the Greenbook fiscal forecasts deteriorated markedly after 1990, with much larger forecast errors for federal government receipts, surpluses, and structural surpluses, particularly at the horizons most relevant for monetary policy. Perhaps surprisingly, this came despite much better forecasts for the unemployment rate and an overall reduction in economic volatility. The asymmetry of risks to the forecast also varied considerably across variables, with the surplus standing out as having unusually large downside risks, apparently due to the asymmetric impact of business cycle shocks.

REFERENCES

Artis, Michael, and Massimiliano Marcellino. 2001. "Fiscal Forecasting: The Track Record of the IMF, OECD, and EC." *Econometrics Journal*, 4: S20–S36.

Auerbach, Alan J. 1994. "The U.S. Fiscal Problem: Where We Are, How We Got Here, and Where We're Going." *NBER Macroeconomics Annual*, 9: 141–175.

Auerbach, Alan J. 1999. "On the Performance and Use of Government Revenue Forecasts." *National Tax Journal*, 52: 767–782.

Auerbach, Alan J., and Yuriy Gorodnichenko. 2012. "Measuring the Output Responses to Fiscal Policy." *American Economic Journal: Economic Policy*, 4(2): 1–27.

Baghestani, Hamid. 2008. "Federal Reserve versus Private Information: Who is the Best Unemployment Rate Predictor?" *Journal of Policy Modeling*, 30: 101–110.

Ball, Laurence, and Dean Croushore. 2003. "Expectations and the Effects of Monetary Policy." *Journal of Money, Credit and Banking*, 35(4): 473–484.

Beetsma, Roel, Massimo Giuliodori, and Peter Wierts. 2009. "Planning to Cheat: EU Fiscal Policy in Real Time." *Economic Policy*, 24(60): 753–804.

Belongia, Michael T. 1988. "Are Economic Forecasts by Government Agencies Biased? Accurate?" *Federal Reseve Bank of St. Louis Review*, 15–23.

Blackley, Paul R., and Larry DeBoer. 1993. "Bias in OMB's Economic Forecasts and Budget Proposals." *Public Choice*, 76(3): 215–232.

Campbell, Bryan, and Eric Ghysels. 1995. "Federal Budget Projections: A Nonparametric Assessment of Bias and Efficiency." *Review of Economics and Statistics*, 77(1): 17–31.

Campbell, Bryan, and Jean-Marie Dufour. 1991. "Over-rejections in Rational Expectations Models: A Nonparametric Approach to the Mankiw-Shapiro Problem." *Economics Letters*, 35: 285–290.

Cimadomo, Jacopo. 2011. "Real-Time Data and Fiscal Policy Analysis: A Survey of the Literature." *European Central Bank Working Paper*, 1408.

Clements, Michael P., Fred Joutz, and Herman O. Stekler. 2007. "An Evaluation of the Forecasts of the Federal Reserve: A Pooled Approach." *Journal of Applied Economics*, 22: 121–136.

Croushore, Dean. 2011. "Frontiers of Real-Time Data Analysis." *Journal of Economic Literature*, 49(1): 72–100.

Croushore, Dean, and Bob Hunt. 2008. "Evaluating Forecasts of Budget Deficits." *Unpublished Working Paper.*

Croushore, Dean, and Simon van Norden. 2014. "Fiscal Policy: Ex Ante and Ex Post." *Federal Reserve Bank of Philadelphia Working Paper 14-22.*

Diebold, Francis X, and Robert S Mariano. 2002. "Comparing predictive accuracy." *Journal of Business & economic statistics*, 20(1).

Elliott, Graham, Ivana Komunjer, and Allan Timmermann. 2005. "Estimation and Testing of Forecast Rationality Under Flexible Loss." *Review of Economic Studies*, 72: 1107–1125.

Frankel, Jeffrey A. 2011. "Over-Optimism in Forecasts by Official Budget Agencies and Its Implications." *National Bureau of Economic Research Working Paper*, 17239.

Harvey, David S, Stephen J Leybourne, and Paul Newbold. 1998. "Tests for forecast encompassing." *Journal of Business & Economic Statistics*, 16(2): 254–259.

Kliesen, Kevin L., and Daniel L. Thornton. 2012. "How Good Are the Government's Deficit and Debt Projections and Should We Care?" *Federal Reseve Bank of St. Louis Review*, 94(1): 21–39.

Leal, Teresa, Javier J. Perez, Mika Tujula, and Jean-Pierre Vidal. 2008. "Fiscal Forecasting: Lessons from the Literature and Challenges." *Fiscal Studies*, 29(3): 347–386.

Mankiw, N. Gregory, and Matthew D. Shapiro. 1986. "Do We Reject Too Often? Small Sample Properties of Tests of Rational Expectations Models." *Economic Letters*, 20: 139–145.

Plesko, George A. 1988. "The Accuracy of Government Forecasts and Budget Projections." *National Tax Journal*, 41: 483–501.

Reischauer, Robert D. 1990. "Taxes and Spending Under Gramm-Rudman-Hollings." *National Tax Journal*, 43: 223–232.

Romer, Christina D., and David H. Romer. 2000. "Federal Reserve Information and the Behavior of Interest Rates." *American Economic Review*, 90(3): 429–457.

Scotese, Carol A. 1995. "Forecast Smoothing and the Optimal Under-Utilization of Information at the Federal Reserve." *Journal of Macroeconomics*, 16(4): 653–670.

Tufte, Edward R. 1983. *The visual display of quantitative information.* Graphics Press.

Strictly Confidential (FR)
Class II FOMC

STAFF PROJECTIONS OF FEDERAL SECTOR ACCOUNTS AND RELATED ITEMS
(Billions of dollars except as noted)

January 29, 1997

Item	Fiscal year[5]				1996				1997				1998			
	1995[a]	1996[a]	1997	1998	Q1[a]	Q2[a]	Q3[a]	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
UNIFIED BUDGET					Not seasonally adjusted											
Receipts[1]	1355	1453	1525	1575	322	446	362	346	341	459	380	363	351	469	392	382
Outlays[1]	1519	1560	1634	1692	393	392	395	405	412	408	410	426	419	424	422	471
Surplus/deficit[1]	-164	-107	-109	-117	-72	54	-33	-59	-71	51	-31	-63	-68	45	-30	-89
On-budget	-226	-174	-184	-187	-84	14	-36	-76	-83	11	-36	-72	-81	2	-36	-101
Off-budget	62	67	75	69	12	39	2	18	12	40	5	8	12	43	6	11
Surplus excluding deposit insurance[2]	-182	-116	-118	-120	-75	52	-34	-65	-72	51	-32	-64	-69	44	-31	-90
Means of financing																
Borrowing	171	130	118	146	80	-23	39	49	58	-31	43	54	69	-25	48	73
Cash decrease	-2	-6	4	0	-1	-16	-6	11	13	-16	-4	10	10	-15	-5	10
Other[3]	-5	-16	-13	-29	-7	-14	0	-1	1	-5	-8	0	-11	-5	-13	6
Cash operating balance, end of period	38	44	40	40	22	38	44	33	20	36	40	30	20	35	40	30
NIPA FEDERAL SECTOR					Seasonally adjusted, annual rate											
Receipts	1459	1544	1625	1684	1523	1576	1582	1625	1606	1626	1645	1663	1672	1691	1711	1731
Expenditures	1629	1683	1749	1798	1678	1702	1703	1730	1741	1755	1772	1782	1792	1802	1815	1829
Consumption expend	455	457	462	463	454	464	461	463	460	461	464	463	464	462	464	460
Defense	304	303	303	305	299	307	305	306	301	302	304	305	306	303	305	303
Nondefense	151	155	159	158	155	156	157	157	159	160	160	157	158	159	159	157
Other expenditures	1175	1226	1287	1335	1225	1239	1241	1267	1281	1294	1308	1319	1328	1340	1351	1369
Current account surplus	-171	-139	-124	-114	-155	-127	-121	-105	-135	-129	-127	-119	-120	-112	-104	-98
Gross investment	65	63	62	62	65	66	64	64	61	62	62	63	62	61	61	60
Current and capital account surplus	-236	-202	-186	-176	-220	-193	-185	-169	-196	-190	-189	-182	-182	-172	-166	-159
FISCAL INDICATORS[4]																
High-employment (HEB) surplus/deficit	-263	-237	-245	-249	-247	-233	-231	-222	-251	-251	-254	-253	-253	-247	-244	-241
Change in HEB, percent of potential GDP	0	- 4	1	1	1	- 2	0	- 1	4	0	0	0	0	- 1	0	0
Fiscal impetus (FI), percent, cal year	-5 6	-1 6	-3 5	-5 5	1 3	1 8	- 2	- 8	-2 2	- 5	1	-2 7	-1 8	-1 3	- 3	-2

1 OMB's July 1996 deficit estimates (assuming the enactment of the President's proposals) are $126 billion in FY97 and $94 billion in FY98 CBO's January 1997 baseline deficit estimates are $124 billion in FY97 and $120 billion in FY98 Budget receipts, outlays, and surplus/deficit include corresponding Social Security (OASDI) categories The OASDI surplus is excluded from the on-budget deficit and shown separately as off-budget, as classified under current law The Postal Service deficit is included in off-budget outlays beginning in FY90

2 OMB's July 1996 deficit estimates (assuming the enactment of the President's proposals), excluding deposit insurance spending, are $134 billion in FY97 and $96 billion in FY98 CBO'S January 1997 baseline deficit estimates, excluding deposit insurance, are $136 billion in FY97 and $124 billion in FY98

3 Other means of financing are checks issued less checks paid, accrued items, and changes in other financial assets and liabilities

4 HEB is the NIPA current and capital account surplus in current dollars, with cyclically sensitive receipts and outlays adjusted to the level of potential output generated by 1 8 percent real growth and an associated unemployment rate of 6 percent Quarterly figures for change in HEB and FI are not at annual rates Change in HEB, as a percent of nominal potential GDP, is reversed in sign FI is the weighted difference of discretionary changes in federal spending and taxes (in chained (1992) dollars), scaled by real federal consumption plus investment For change in HEB and FI, negative values indicate restraint

5 Fiscal 1995 data for the unified budget come from OMB,' fiscal 1996 and quarterly data come from the Monthly Treasury Statement and may not sum to OMB fiscal year totals

a--Actual

FIGURE 1. A SAMPLE GREENBOOK PAGE

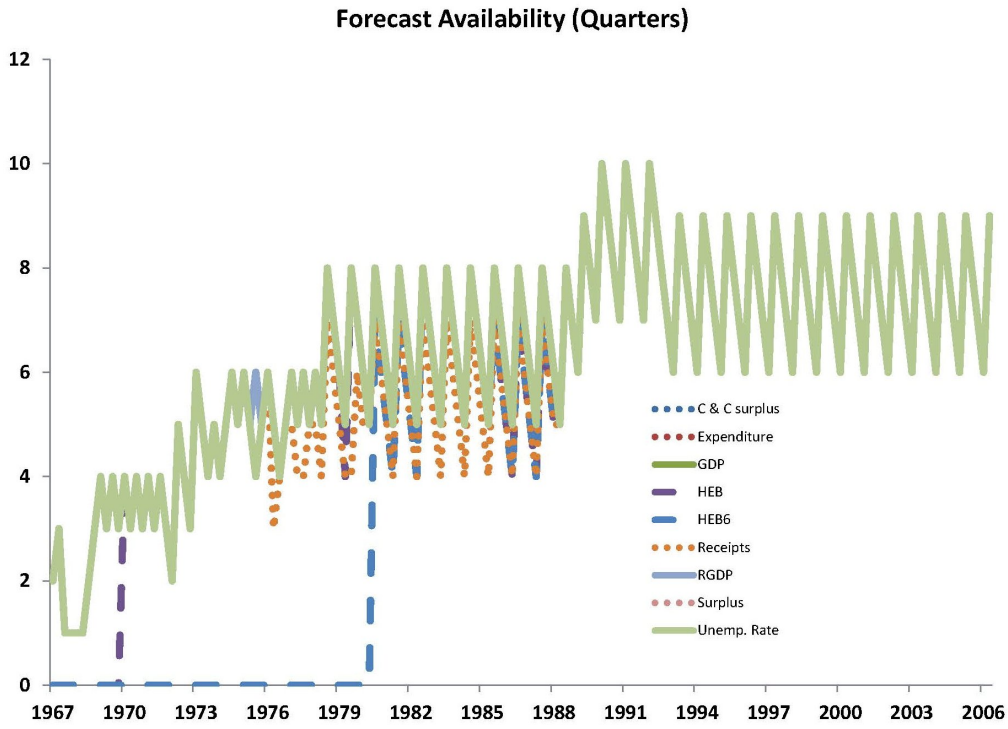

FIGURE 2. GREENBOOK FORECAST HORIZONS BY DATE AND SERIES

Note: Counts are from the first FOMC meeting of each quarter.

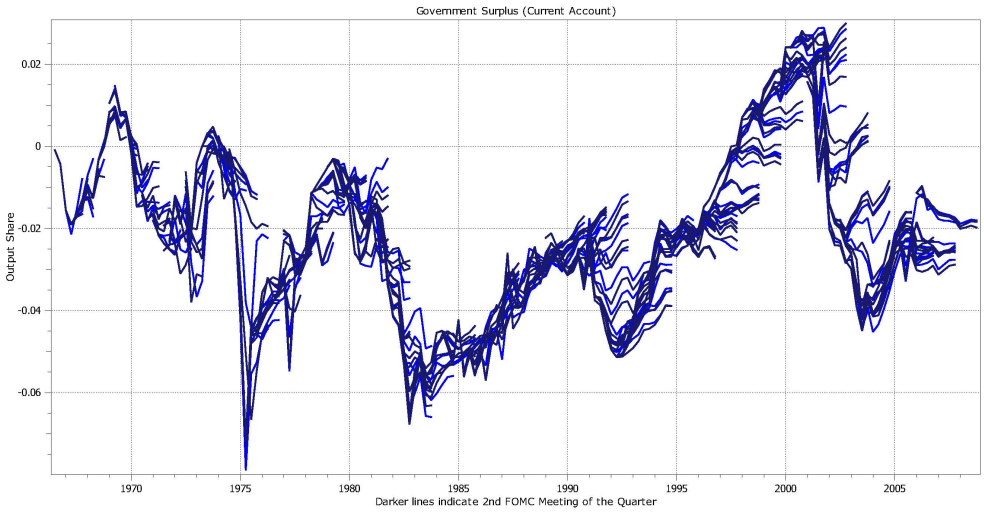

FIGURE 3. GREENBOOK GOVERNMENT SURPLUS FORECASTS

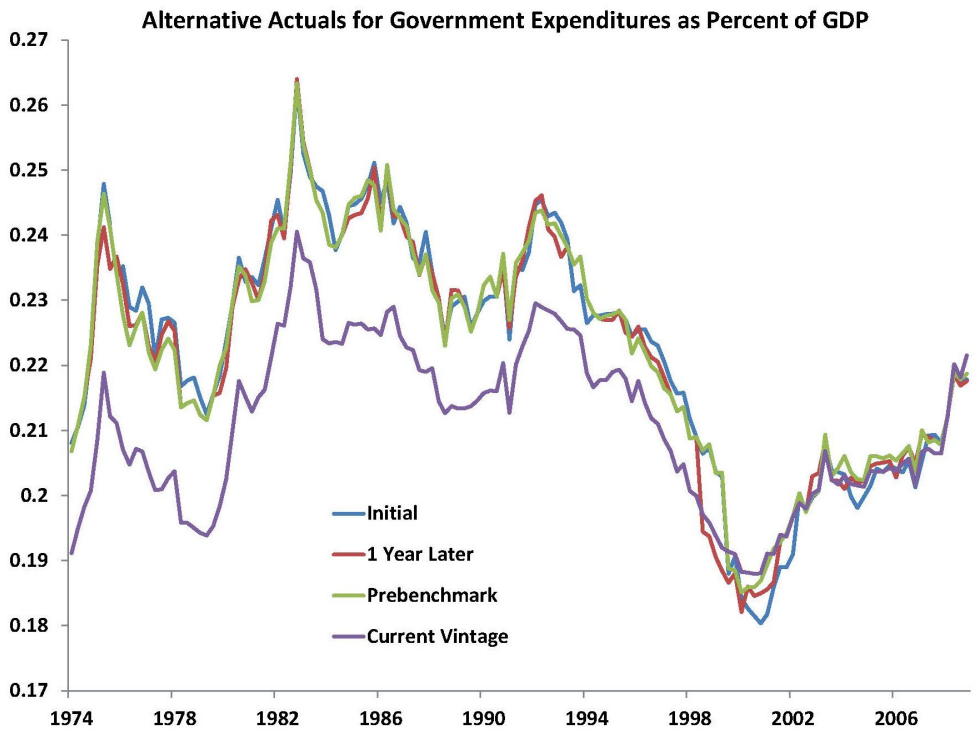

FIGURE 4. REALIZED VALUES OF GOVERNMENT EXPENDITURES BASED ON ALTERNATIVE CONCEPTS

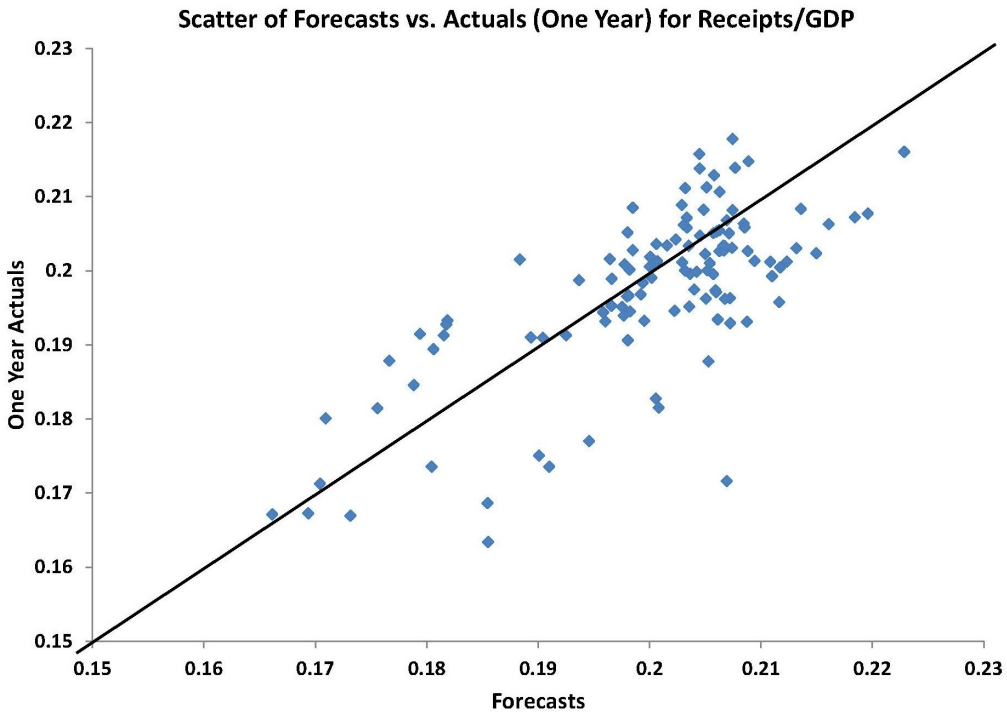

Figure 5. Scatterplot of Four-Quarter-Ahead Receipt Forecasts Against Realized Values

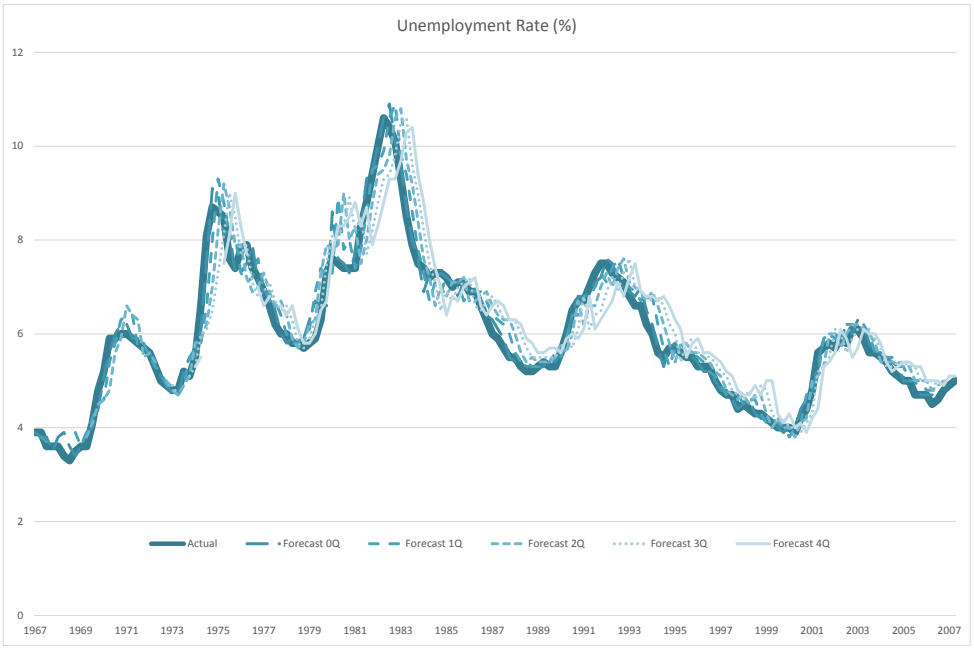

FIGURE 6. UNEMPLOYMENT RATE FORECASTS AND OUTCOMES

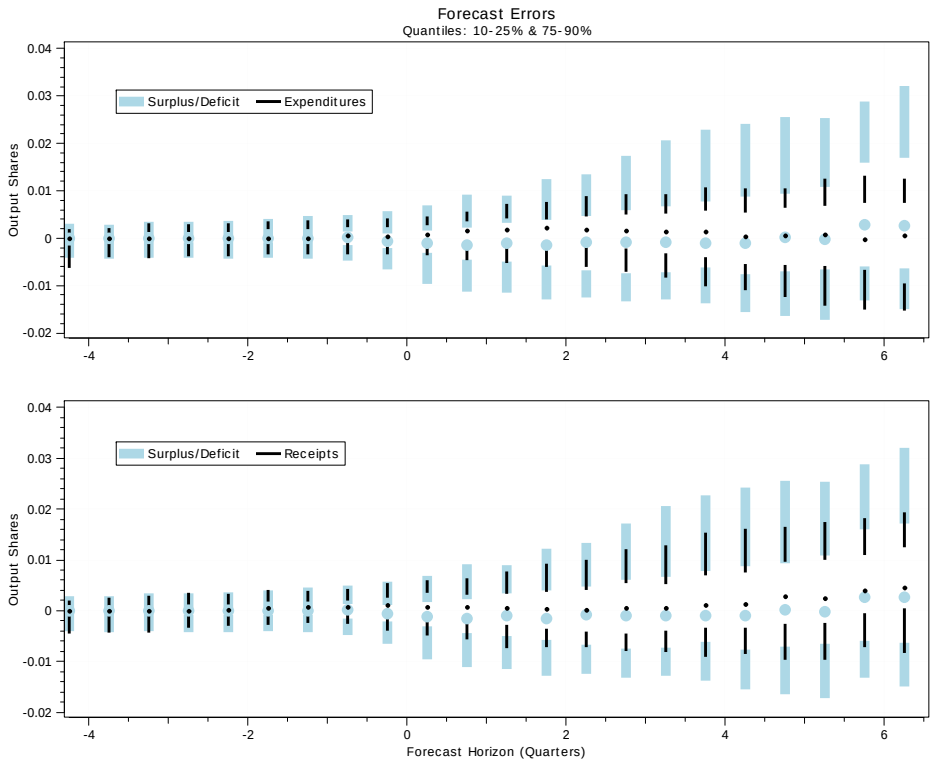

FIGURE 7. FORECAST ERRORS FOR SURPLUS, RECEIPTS AND EXPENDITURES

Note: The simplified box plots above compare five forecast error quantiles for the Surplus, Receipts and Expenditures at each forecast horizon. In each case, the upper line runs from the 90th to the 75th percentiles while the lower line runs from the 25th to the 10th. A dot indicates the median.

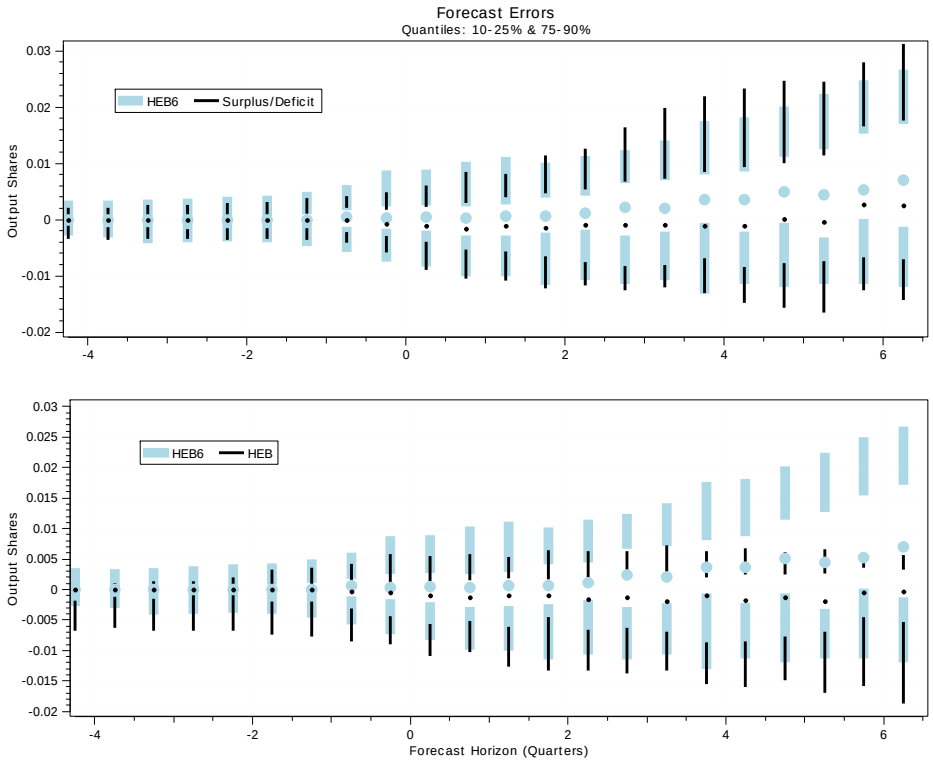

FIGURE 8. FORECAST ERRORS FOR THREE MEASURES OF SURPLUS

Note: The simplified box plots above compare five forecast error quantiles for the Surplus and two measures of the structural surplus (HEB and HEB6) at each forecast horizon. In each case, the upper line runs from the 90th to the 75th percentiles while the lower line runs from the 25th to the 10th. A dot indicates the median.

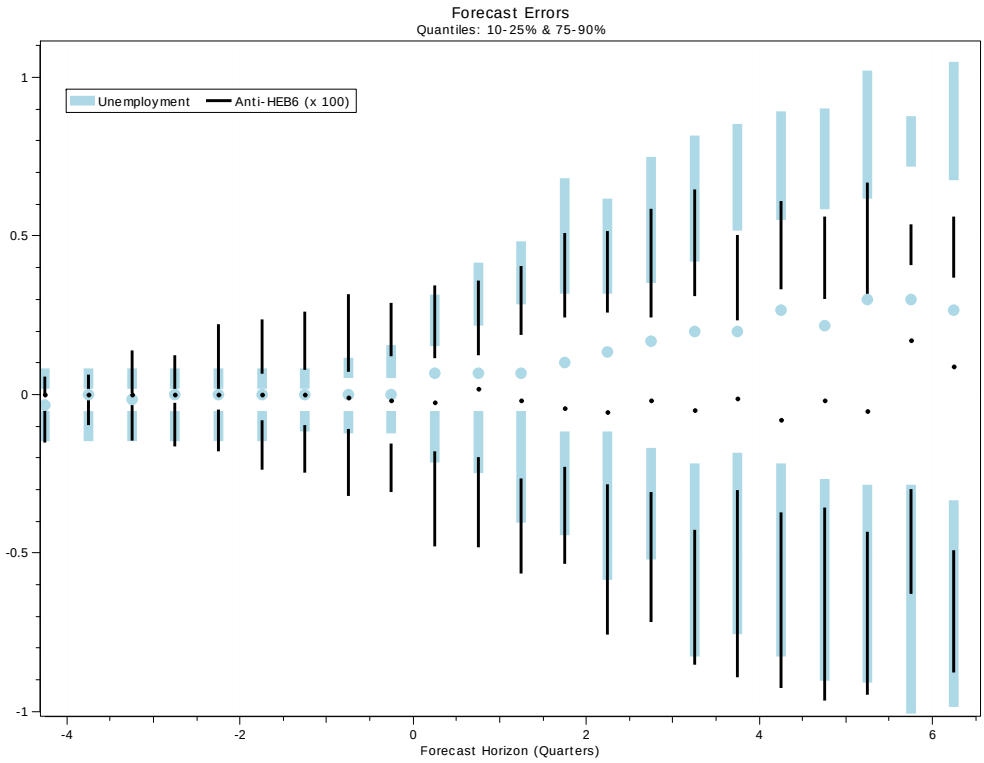

FIGURE 9. FORECAST ERRORS FOR UNEMPLOYMENT AND THE CYCLICAL SURPLUS.

Note: The simplified box plots above compare five forecast error quantiles for the unemployment rate and the Cyclical Surplus (Anti-HEB: calculated as the Surplus minus HEB6) at each forecast horizon. In each case, the upper line runs from the 90th to the 75th percentiles while the lower line runs from the 25th to the 10th. A dot indicates the median.