WORKING PAPER NO. 17-13
FISCAL SURPRISES AT THE FOMC

Dean Croushore
Professor of Economics and Rigsby Fellow,
University of Richmond and
Visiting Scholar, Federal Reserve Bank of Philadelphia

Simon van Norden
HEC Montréal, CIRANO, and
Visiting Scholar, Federal Reserve Bank of Philadelphia

June 2017

RESEARCH DEPARTMENT, FEDERAL RESERVE BANK OF PHILADELPHIA
Ten Independence Mall, Philadelphia, PA 19106-1574 • www.philadelphiafed.org/research-and-data/

Fiscal Surprises at the FOMC

By Dean Croushore and Simon van Norden*

This paper provides a detailed examination of a new set of fiscal forecasts for the U.S. assembled by Croushore and van Norden (2017) from FOMC briefing books. The data are of particular interest because (1) they afford a look at fiscal forecasts over six complete business cycles and several fiscal policy regimes, covering both peacetime and several wars, (2) the forecasts were precisely those presented to monetary policymakers, (3) they include frequently updated estimates of both actual and cyclically adjusted deficits, (4) unlike most other U.S. fiscal forecasts, they were neither partisan nor constrained by unrealistic assumptions about future fiscal policy, and (5) forecasts for other variables (GDP growth, inflation) from the same forecasters are known to compare favorably with most other available forecasts.

We detail the performance of forecast federal expenditures, revenues, surpluses, and structural surpluses in terms of accuracy, bias, and efficiency. We find that (1) fiscal forecast errors can be economically large, even at relatively short forecast horizons, (2) while the accuracy of unemployment rate forecast errors improved after 1990, that of most fiscal variables deteriorated considerably, (3) there is limited evidence of forecast bias, and most of this evidence is confined to the period before 1993, (4) the forecasts appear to be efficient with respect to both the fed funds rate and CBO projections, and (5) cyclically adjusted deficit forecasts appear to be over-optimistic around both business cycle peaks and troughs.

JEL: E62, H68

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

Research on monetary policy has focused on rules (such as the Taylor rule) and the evaluation of forecasts (such as those by the Federal Reserve Board's staff in the FOMC Greenbook). While fiscal policy has gained renewed attention in the aftermath of the 2008 financial crisis, the corresponding literature on fiscal policy rules and the quality of fiscal forecasts is much more sparse.[1] Much of the literature on forecasts of U.S. fiscal policy (as we discuss below) analyzes U.S. Congressional Budget Office (CBO) forecasts. But the CBO is required *by law* to produce forecasts under the assumption of no changes in tax policy or spending policy over the forecast horizon. For this reason, other forecasts are likely to be more realistic predictors of fiscal policy and better measures of expected fiscal policy. Croushore and van Norden (2017) argue that some of the best work on fiscal policy in recent years has been done on Eurozone data, due in part to the availability of suitable data sets. Their paper begins to remedy that situation by documenting a new coherent database of high-quality forecasts of U.S. federal fiscal policy variables. We build on their work by providing new in-depth analysis of their forecast performance.

Forecast 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. In particular, 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 there-

* 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. Both authors would like to thank CIREQ and the Real-Time Data Research Center of the Federal Reserve Bank of Philadelphia for their hospitality. The views expressed in this paper are those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of Philadelphia or the Federal Reserve System. This paper is available free of charge at *www.philadelphiafed.org/research-and-data/publications/working-papers*.

[1] Bernanke (2017) describes the importance of fiscal policy for monetary policy formulation, noting "Fiscal policy influences the economy through many channels. ... To project the impact of a proposed fiscal package on the economy, Fed modelers and policymakers must assess the size and timing of these demand and supply effects, which they do based both on theory and historical experience. ... When I was Fed chair, I argued on a number of occasions against fiscal austerity (tax increases, spending cuts). ... I pushed (unsuccessfully) for fiscal policies to increase aggregate demand and job creation." Croushore and van Norden (2017) survey historical testimony on the role of fiscal policy in the formulation of U.S. monetary policy.

fore follow Croushore and van Norden (2017) and carefully match fiscal forecasts with contemporaneous data vintages of other key variables to allow us to properly understand the information available to policymakers.

We begin the paper in section I, which describes the Croushore and van Norden (2017) Greenbook data set. Section II characterizes how the qualitative behavior of the forecast errors varies across variables and forecast horizons. In section III we test for both unconditional as well as conditional forecast bias. Section IV tests whether the forecasts are efficient, or whether forecast errors can be predicted with the aid of other series, such as past forecast errors, other available forecasts, or monetary policy instruments. We summarize the results and draw conclusions in section V.

I. Greenbook Data

This section provides a brief introduction to the Croushore and van Norden (2017) Greenbook data set.[2] The Greenbook is a summary of economic conditions, trends, and forecasts prepared for every meeting of the FOMC. It was first prepared for the July 1966 FOMC meeting and the last we included was for the December 2010 meeting, covering 419 meetings of the FOMC over 44 years.[3] Most of our fiscal variables (Surplus, Revenues and Expenditures) first appeared in the Greenbook dated 9 August 1967 while the first appearance of the High-Employment Budget Surplus/Deficit (HEB) variable was dated 29 April 1970.

Croushore and van Norden (2017) collected all Greenbook estimates for their selected series. This included estimates for future periods (forecasts), current periods (nowcasts) and historical periods (backcasts). We collectively refer to all of these as *forecasts* although some prefer the term "projection" to emphasize the

[2]The web appendix to Croushore and van Norden (2017) provides additional detail. The data set is available via that paper's replication files.

[3]Greenbooks are not publicly available for at least five years after their creation. In June 2010, the Greenbook was merged with the Bluebook (a discussion of policy options) to form the Tealbook. As this is near the end of our sample period, we continue to use the term Greenbook to mean the Greenbook prior to June 2010 and the Tealbook after that.

conditional nature of these estimates. FOMC meeting dates are slightly irregular, but for most of the period there were exactly two meetings per quarter. To standardize the forecast horizons that we examine, we follow Croushore and van Norden (2017) and restrict our analysis to the vintages from the (F)irst and the (L)ast FOMC meeting of each quarter. The available forecast horizons sometimes varied across variables. The number of observations and the forecast horizons included in each series also varied considerably over time. In some of the earliest vintages, series might not contain more than 5 quarters of historical estimates and forecasts, whereas later vintages could contain up to 20 quarters. As we increase the forecast horizons beyond 6 quarters, the number of available forecasts begins to drop sharply and their coverage becomes both more sporadic and heavily weighted toward the later portion of our sample period. For comparability across horizons, we therefore at times exclude the longer horizons from our analysis.

Of course, forecast evaluation requires a measure of observed 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. We examined a variety of different "outcome" concepts to provide alternative characterizations of forecast performance. They are

Initial Release: This is the initial quarterly estimate published by the responsible official statistical agency (BEA or BLS).

One Year: This is the official quarterly estimate that was available precisely one year after the publication of the First Release. For example, if the First Release was published on September 23, 1998, and revisions were published on August 26, 1999, and September 29, 1999, the August 1999 estimate would be the One Year estimate. This typically incorporates the first annual revision common to most official series.

Last Greenbook: This is the last value recorded in the Greenbook, typically one or more years after the quarter to which it refers. This is primarily important

as a measure for HEB, which has no counterpart in official statistics; it is only calculated by Board staff.

Pre-Benchmark: This is the last official estimate reported prior to a benchmark revision of the series. This is intended to capture the most precise available estimate of the same concept that the staff were forecasting and has previously been used in the literature to measure data revisions.[4] We discuss the identification and importance of benchmark revisions in the Appendix.

Final: This is a "contemporary" estimate, which in our case was the official estimate as of May 2016.

These revisions may reflect the incorporation of new information as published preliminary estimates are refined in the quarters immediately following their initial publication. It may also reflect conceptual changes in the definition of the series, such as the change from GNP to GDP or from a fiscal surplus to a fiscal current account surplus. We refer to the latter as "benchmark" revisions. Each of our series were affected, to greater or lesser degrees, by benchmark changes. In our results below, we focus on results which use the Pre-Benchmark estimate although our conclusions were typically robust to this choice. (We note exceptions below.) In using the Pre-Benchmark estimates, we omit forecasts made just before a benchmark change for which official estimates were published only after the change.

A. Variables

GNP and GDP: Our outcome measures for these series were taken from ALFRED series *GNP* and *GDP.* Our primary use of these series is to express various fiscal series as a fraction of the overall size of the U.S. economy.

Receipts, Expenditures and Surplus/Deficit: Outcomes for the Surplus/Deficit were measured by ALFRED series *FGDEF: Net Federal Government Saving.* Outcomes for Receipts were taken from *FGRECPT: Federal Government*

[4]For example, see Aruoba (2008) or Croushore and van Norden (2017).

Current Receipts, and for Expenditures from *FGEXPND: Federal Government: Current Expenditures.*

HEB: The *High-Employment Budget Surplus/Deficit* (HEB) is the Greenbook's estimate of a cyclically adjusted or "structural" budget deficit. This is the Board staff's counterfactual estimate of what the surplus (or deficit) would be if the unemployment rate were at a constant reference level over the forecast horizon. The budget deficit concept used in HEB always corresponds to that used in the Surplus/Deficit measure; prior to 1996 this was the overall Surplus or Deficit, and this was replaced by the Government Current Account Surplus/Deficit thereafter. No statistical agency publishes estimates for our structural deficit measure, HEB; we therefore just compare its forecasts with the last reported value (*Last*).

The level of unemployment used to calculate HEB is not always explicitly mentioned, but drifted upward (from near 4.0% in the earliest part of our sample) before major changes were introduced in 1980. From November 1980 until March 1983, two alternative HEB estimates were presented, based on a 6.1% and a 5.1% reference level of unemployment. From May 1983 until August 1983 these were replaced by rates of 6.0% and 5.0%. Thereafter, the reference level was constant at 6.0%. We assume that these changes reflected uncertainty and disagreement within the Board about the natural rate of unemployment. The table design during the "dual-rate" period gave greater prominence to the 6.1% (and then the 6.0%) reference level.

We found that the revision of the reference level of unemployment appeared to have a qualitatively important effect on the HEB estimates. We therefore consider two different sets of HEB estimates; the full series as well as the subset (HEB6) which only considers estimates based on a 6.0% or 6.1% reference level. We make no attempt to adjust the HEB6 series for the change from 6.1% to 6.0%. We also calculate the difference between the HEB (and HEB6) estimates and the overall Surplus/Deficit estimates as the Board Staff's implied estimate of the cyclical Surplus/Deficit.

Unemployment: Outcomes for this series were measured by ALFRED series *UNRATE: the Civilian Unemployment Rate.* Greenbooks only report the unemployment rate to one decimal place. Starting with the official estimate published on 9 February 1967, the labor force was redefined to count only those age 16 and over instead of 14 and over. This never caused revisions of more than 0.1% in absolute value in our data set. There were no benchmark revisions to unemployment after that date. We therefore chose to ignore benchmark revisions in the unemployment rate and do not use a "Pre-Benchmark" measure of outcomes.

II. Forecast Accuracy

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 examine the size of their forecast errors and how they vary by horizon and variable.

Our first step is to 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 that they attempt to predict.[5] As values approach one, however, the forecasts capture less and less of the variation in the target variable. Table 1 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 date recedes into the future, we expect to see a steady rise in the relative variance of the forecast errors. Results for our full 1967–2010 sample are shown in Figure 1.

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. By far the most accurate forecasts were those

[5]The realization series used is Pre-benchmark for Expenditures, Receipt s, and Surplus, Last Greenbook for HEB and HEB6, and Final for Unemployment.

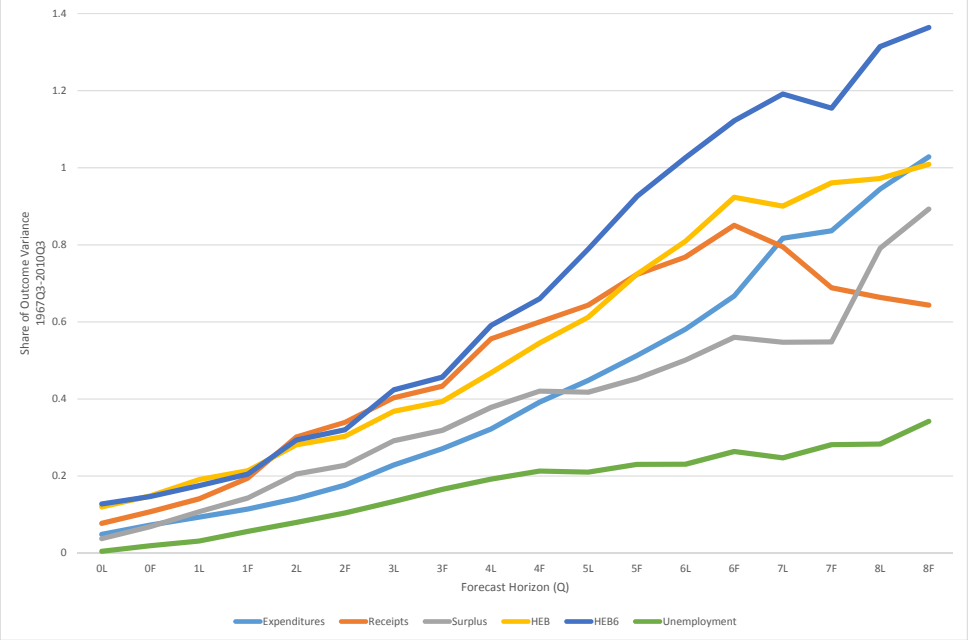

Figure 1. Relative Variance of Forecast Errors: 1967Q3–2010Q4

for unemployment rate, where forecasts at an 8Q horizon still captured two-thirds of the unconditional variance of unemployment. Forecast accuracy for the fiscal variables declined much more rapidly, however, with forecast errors for the Surplus reaching 40 % by the 4Q, 50% by 6Q and 80% by the 8Q horizon. At all but the longest horizons, forecast errors for Receipts were worse, with error variances already exceeding 80% by 6Q. Forecasts of structural deficits, HEB and HEB6 were generally the worst performers with error variances greater than 50% after only 4Q and eventually reaching or exceeding 100%. (Values greater than 100% imply that one would be better off replacing the forecast with the unconditional mean of the target variable.) To the extent that monetary policy was sensitive to

these fiscal forecasts, they could have been an important source of policy errors.[6]

These results conceal some important variations across time, however, as we show in Table 1 where we distinguish between forecast errors prior to 1991 and those after 1990.[7] The latter period is dominated by the Great Moderation, and the last column of the table shows that unemployment forecasts were more accurate, with forecast error variances at the 4Q horizon much less than half the size of the earlier period.

Despite the general decline in economic volatility during the latter period, the forecast performance for the fiscal variables deteriorated noticeably. Comparing the lower panel of the table with the upper, we see that differences in the magnitude of forecast errors for the Surplus were minor until the 3Q horizon, but then quickly rose with the forecast horizon to be more than double by 5Q. This largely mirrors the behavior of Expenditure and HEB forecasts, while the deterioration in forecasts of Receipts is already marked by 2Q. The most dramatic change, however, is in HEB6, with forecast error variances at least double those of the earlier period at every horizon from 1Q onwards, and sometimes an order of magnitude larger. It therefore seems that while overall economic activity was less volatile after 1990, fiscal policy became harder to forecast.[8]

An examination of the forecast errors showed 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

[6]Croushore and van Norden (2017) find that the Fed Funds Rate Target was significantly influenced by the Board's fiscal projections.

[7]Note that forecast error variances are still expressed here relative to the full-sample variances of the variable forecast. This allows a direct comparison of the figures in the upper and lower portions of the table.

We explored a small number of alternative sample periods. Qualitatively, results were similar regardless of whether the early sample started in 1967Q3 or 1974Q4, and whether the later sample ended in 2006Q4 or 2010Q4. Changing the break date from 1990Q4 to 1985Q4 to better capture the Great Moderation somewhat reduced the differences between the samples, suggesting that the 1986–1990 period was more similar to the earlier sample.

[8]For comparison, Table 12 in the Appendix provides results excluding the post-2006 period.

Table 1—Forecast Error Variances

Horizon	Expenditures	Receipts	Surplus	HEB	HEB6	Unemployment
1967Q3–1990Q4						
0L	0.038	0.059	0.036	0.128	0.086	0.006
0F	0.073	0.106	0.078	0.166	0.089	0.027
1L	0.105	0.136	0.122	0.219	0.093	0.045
1F	0.141	0.198	0.177	0.230	0.085	0.080
2L	0.159	0.269	0.221	0.264	0.071	0.116
2F	0.192	0.291	0.246	0.276	0.068	0.151
3L	0.227	0.286	0.278	0.308	0.087	0.194
3F	0.275	0.292	0.303	0.310	0.088	0.236
4L	0.274	0.361	0.306	0.335	0.056	0.271
4F	0.340	0.330	0.332	0.406	0.084	0.299
5L	0.308	0.219	0.180	0.411	0.090	0.299
5F	0.323	0.125	0.141	0.498	0.229	0.330
6L	0.287	0.125	0.122	0.687	0.303	0.336
6F	0.342	0.099	0.164	0.868	0.326	0.377
1991Q1–2010Q4						
0L	0.049	0.095	0.040	0.111	0.151	0.003
0F	0.061	0.108	0.053	0.131	0.178	0.009
1L	0.066	0.148	0.086	0.164	0.222	0.012
1F	0.074	0.196	0.096	0.201	0.271	0.024
2L	0.115	0.350	0.188	0.305	0.412	0.029
2F	0.156	0.409	0.212	0.336	0.454	0.042
3L	0.233	0.560	0.325	0.450	0.608	0.051
3F	0.276	0.597	0.357	0.487	0.659	0.078
4L	0.372	0.763	0.497	0.624	0.843	0.091
4F	0.435	0.818	0.545	0.692	0.936	0.122
5L	0.532	0.923	0.662	0.790	1.068	0.128
5F	0.613	1.007	0.726	0.894	1.209	0.158
6L	0.693	1.017	0.745	0.915	1.238	0.165
6F	0.790	1.063	0.808	1.030	1.393	0.196
7L	0.945	0.946	0.723	1.033	1.397	0.195
7F	0.970	0.828	0.696	1.013	1.370	0.222
8L	1.139	0.818	0.856	1.109	1.499	0.231
8F	1.225	0.834	0.836	1.107	1.497	0.280

Note: Forecast error variances are shown as a fraction of the unconditional variance of the underlying series over the period 1967Q3–2010Q4. Because HEB and HEB6 are identical in the post-1990 period, their figures in the lower portion of the Table differ only to the extent that their variances differed in the pre-1991 sample.

Forecasts are taken from the first FOMC meeting in 1967Q3 until the last meeting in 2010Q4. The "Horizon" column shows a number denoting the forecast horizon in quarters and the letter L or F, which stand for the (F)irst or (L)ast FOMC meeting of the quarter.

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.

Forecasts with horizons longer than 6Q were not available for all series prior to 1991.

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.

A. The Distribution of Forecast Errors

Another way to compare forecast performance is to compare the distribution of forecast errors across forecast horizons and across variables. This is succinctly summarized in Figures 2 and 3. Each figure shows simplified box plots (due to Tufte (1983)) describing the distribution of forecast errors for each forecast horizon. At each forecast horizon, vertical lines link the 90th to the 75th percentiles as well as the 25th to the 10th percentile. Dots indicate the median, maximum, and minimum forecast errors. By overlaying box plots for two series, we see how their distributions compare and vary with the forecast horizon.[9]

Figure 2 compares the distribution of forecast errors for the Surplus with that of Expenditures (upper panel) and that of Receipts (lower panel). Note that substantial forecast errors are common; the 80% confidence intervals shown for the Surplus have a width rising from roughly 1% for the last nowcast of the quarter to just over 3% at the 4Q horizon and roughly 5% at the 6Q horizon. Of course, the largest forecast errors are more dramatic still: for example, in both 1975Q2 and 2009Q2, the size of the deficit was under-predicted by more than 5% of GDP in forecasts made only 2Q previously. This general inaccuracy reflects forecast errors in both revenues and expenditures, with similar degrees of forecast accuracy in both series across most forecast horizons.

We also see that forecast errors are clearly skewed. While median forecast

[9]Forecast errors are based on current vintage outcomes for the Unemployment Rate, on Last Greenbook values for HEB6, 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.

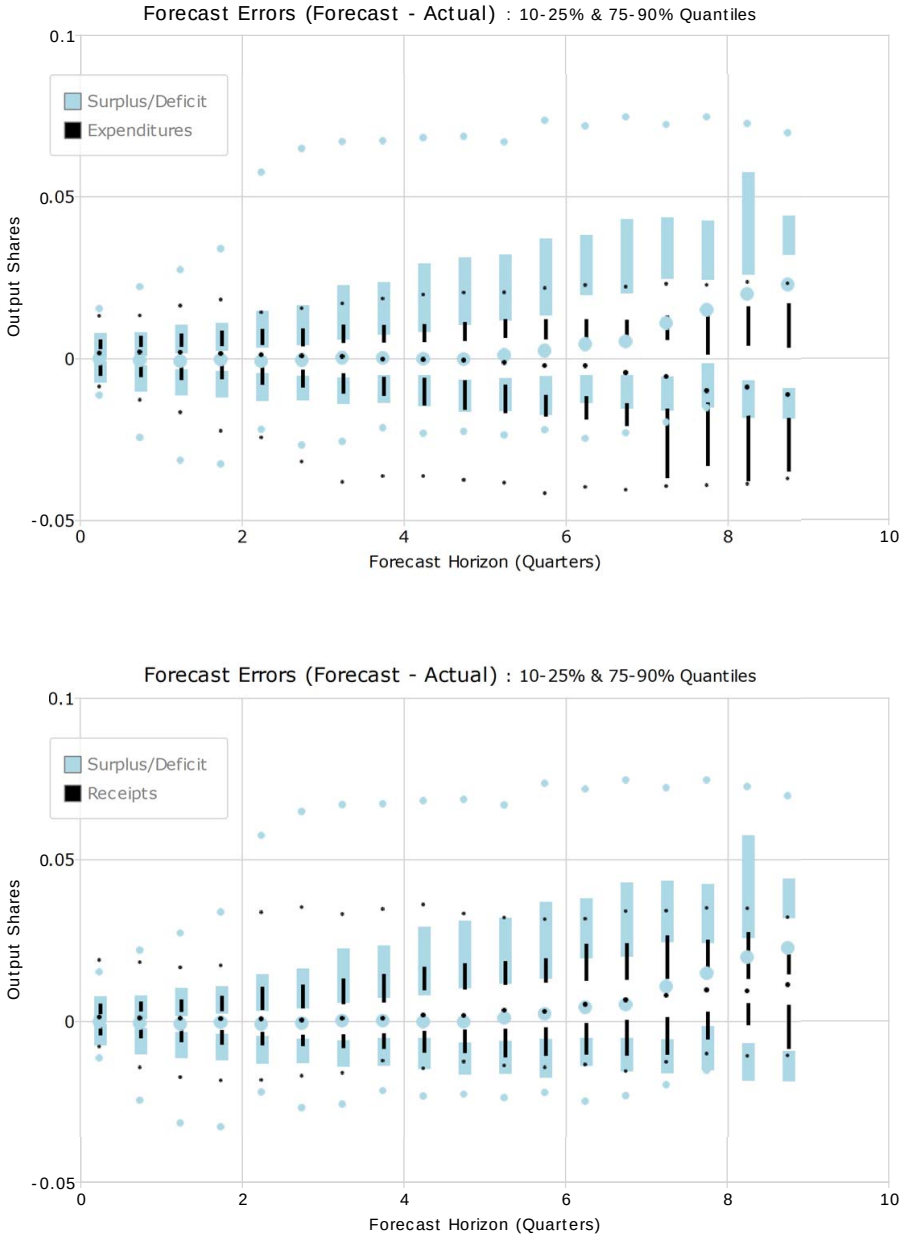

Figure 2. Forecast Errors for Surplus/Deficit, Expenditures, and Receipts

Note: The simplified box plots above compare forecast error quantiles for the Surplus/Deficit, Expenditures and Receipts. At each forecast horizon, vertical lines link the 90th to the 75th percentiles as well as the 25th to the 10th percentile. Dots indicate the median, maximum, and minimum forecast errors.

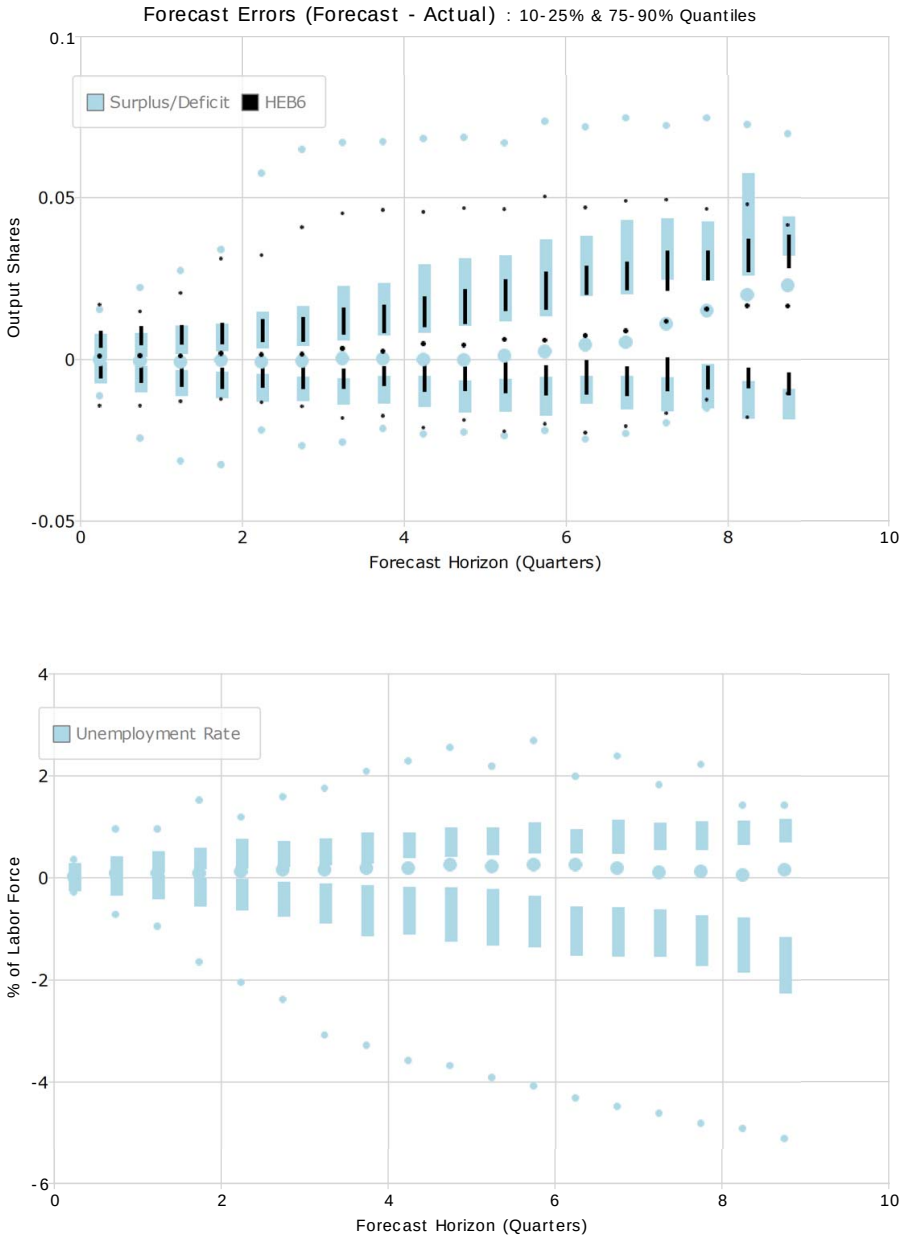

FIGURE 3. FORECAST ERRORS FOR SURPLUS/DEFICIT, HEB6 AND UNEMPLOYMENT

Note: The simplified box plots above compare forecast error quantiles for the Surplus/Deficit, HEB6 and the Unemployment Rate. At each forecast horizon, vertical lines link the 90th to the 75th percentiles as well as the 25th to the 10th percentile. Dots indicate the median, maximum, and minimum forecast errors.

errors are close to zero from the nowcasts until the 4Q forecasts, thereafter they diverge in the direction of overly optimistic forecasts (high surplus and receipts, low deficits and expenditures) with median errors exceeding 2% for the Surplus at the longest horizons. However, the interquartile range and the 80% confidence intervals show clear evidence of asymmetric risks at nearly all horizons, with large negative outcomes (higher than forecast expenditures and deficits, lower than forecast revenues and surpluses) more likely than positive ones.

The upper panel of Figure 3 compares the same forecast errors for the surplus seen in Figure 2 with those of the structural surplus HEB6.[10] We see that the forecast errors for HEB6 are almost always less widely dispersed than those for the Surplus, presumably reflecting the impact of the additional business cycle uncertainty on the Surplus. However, we see the same positive drift in median forecast errors at longer horizons in HEB6 that we saw above in the Surplus, with median errors eventually exceeding 1.5% of GDP. We also find important skewness in the distribution at almost all forecast horizons.

While HEB and HEB6 attempt to correct for the direct effects of business cycle fluctuations, they should capture the responsiveness of fiscal policy to such shocks. The asymmetry in the forecast errors for HEB6 would therefore be consistent with a countercyclical fiscal policy that responds to asymmetric business cycle shocks. The lower panel of Figure 3 shows the distributions of the unemployment rate forecast errors, which confirms the asymmetric nature of the business cycle shocks in our sample. Unlike the Surplus and HEB6, median forecast errors for unemployment are close to zero at all horizons.

III. Forecast Bias

When testing for forecast bias, tests of forecasts covering horizons longer than the frequency of the observations are subject to the standard overlapping obser-

[10]Figure 10 in the Appendix shows that the distributions of forecast errors for HEB and HEB6 were quite similar. For brevity, therefore, we limit our discussion here to HEB6.

vations problem.[11] 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 2. The table shows p-values for the null hypothesis of no bias for five different forecast horizons, four different measures of outcome concepts (Last, Initial, One Year, and Prebenchmark), two different meeting times during the quarter and six different variables (Surplus, Expenditures, Receipts, HEB, HEB6, and the Unemployment Rate.)[12]

Overall, it is remarkable that the evidence of bias is generally robust to the measure of outcomes we use. 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 (expenditures were less than forecast, on average) at a zero-quarter horizon, but not for longer horizons. The evidence for forecasts of receipts is mixed, with weaker evidence of bias at the shortest horizon but stronger evidence at longer horizons (receipts were less than forecast, on average). Benchmark revisions appear to account for much of this, however, as the evidence of bias is much weaker when using the Prebenchmark measure of outcomes. HEB forecasts are biased for all horizons (structural surpluses were less than forecast, 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.[13] At longer horizons, evidence of bias is marginal.

To understand why the Receipt forecasts might be biased, we plot the four-

[11] 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.

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

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

TABLE 2—SUMMARY RESULTS OF BIAS TESTS

		Surplus		Expenditures		Receipts	
Horizon	Concept	First	Last	First	Last	First	Last
0	Last	0.08*	< 0.01***	0.06*	0.13	< 0.01***	< 0.01***
	Initial	0.57	0.64	0.02**	0.03**	0.20	0.03**
	One Year	0.75	0.16	< 0.01***	< 0.01***	< 0.01***	< 0.01***
	Prebenchmark	0.48	0.87	0.06*	0.13	0.31	0.07*
1	Last	0.20	0.16	0.51	0.48	0.07*	0.07*
	Initial	0.95	0.84	0.39	0.29	0.50	0.64
	One Year	0.62	0.64	0.09*	0.06*	0.02**	0.02**
	Prebenchmark	0.96	0.86	0.53	0.51	0.59	0.74
2	Last	0.09*	0.09*	0.82	0.96	0.04**	0.05*
	Initial	0.37	0.41	0.71	0.86	0.27	0.28
	One Year	0.23	0.25	0.64	0.46	0.02**	0.02**
	Prebenchmark	0.39	0.43	0.78	0.92	0.21	0.22
3	Last	0.06*	0.07*	0.60	0.80	0.02**	0.02**
	Initial	0.16	0.22	0.58	0.72	0.07*	0.09*
	One Year	0.11	0.15	0.97	0.77	< 0.01***	< 0.01***
	Prebenchmark	0.17	0.24	0.53	0.73	0.07*	0.08*
4	Last	0.06*	0.06*	0.54	0.64	< 0.01***	< 0.01***
	Initial	0.13	0.14	0.62	0.69	0.04**	0.04**
	One Year	0.09*	0.10*	0.81	0.96	< 0.01***	< 0.01***
	Prebenchmark	0.14	0.15	0.48	0.58	0.04**	0.04**
1-4	Last	0.12	0.09*	0.79	0.83	0.03**	0.04**
	Initial	0.37	0.33	0.86	0.82	0.16	0.19
	One Year	0.20	0.16	0.83	0.72	< 0.01***	< 0.01***
	Prebenchmark	0.32	0.29	0.61	0.63	0.14	0.16

		HEB		HEB6		Unemployment	
Horizon	Concept	First	Last	First	Last	First	Last
0	Last	< 0.01***	< 0.01***	0.04**	0.03**	0.01**	0.13
	Initial					0.03**	0.52
	One Year					0.03**	0.52
1	Last	< 0.01***	< 0.01***	0.05*	0.13	0.12	0.03**
	Initial					0.18	0.07*
	One Year					0.18	0.07*
2	Last	< 0.01***	< 0.01***	0.04**	0.07*	0.47	0.29
	Initial					0.54	0.37
	One Year					0.54	0.37
3	Last	< 0.01***	< 0.01***	0.02**	0.03**	0.78	0.58
	Initial					0.83	0.65
	One Year					0.83	0.65
4	Last	< 0.01***	< 0.01***	0.02**	0.02**	0.85	0.74
	Initial					0.89	0.78
	One Year					0.89	0.78
1-4	Last	< 0.01***	< 0.01***	0.09*	0.08*	0.55	0.41
	Initial					0.60	0.47
	One Year					0.60	0.47

Note: The figures shown are p-values for tests of the null hypothesis that the mean forecast error is zero. 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 %). Calculations use Newey-West heteroscedasticity and autocorrelation robust standard errors with the number of lags equal to the forecast horizon minus one. The sample period is 1974Q4 to 2010Q4, except for HEB6, for which the sample begins in 1981Q1. *First* and *Last* refer to the first and last FOMC meetings of each quarter.

quarter-ahead forecast against the one-year realized value in Figure 4. 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. The forecast errors in receipts were particularly large in the late 1990s and early 2000s, when the Greenbook persistently forecast a rise in receipts that did not materialize. In this period, the Greenbook (and other forecasters) apparently 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.

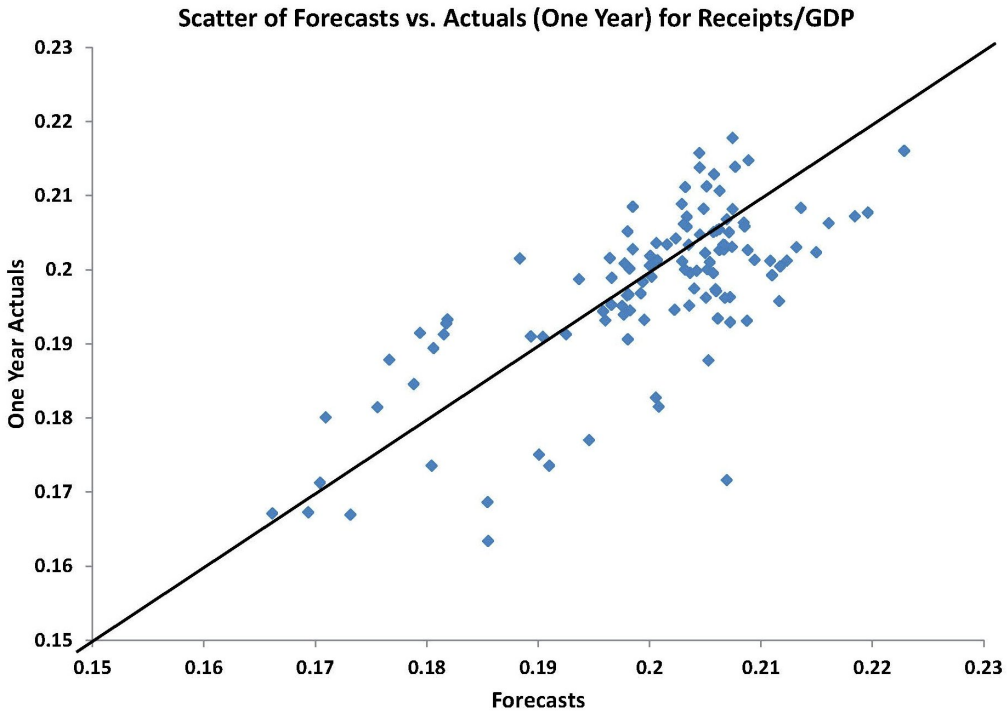

Figure 4. Scatterplot of Four-Quarter-Ahead Receipt Forecasts Against Realized Values (The black diagonal indicates the 45-degree line.)

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 receipts 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 receipts, 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 (i.e., overly pessimistic) at longer horizons.[14]

Others have previously tested for bias in the Greenbook unemployment rate forecasts, but their results have been mixed. Baghestani (2008) uses forecasts made from 1983 to 2000 and similarly finds significant evidence that the Board staff tended to overpredict the unemployment rate, and that the bias increased with the forecast horizon. In contrast, Clements, Joutz and Stekler (2007) used 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. Alternatively, it is possible that the inclusion of the post-2007 period may have changed the evidence of bias.

This led us to examine whether our results were consistent over time. We did so by splitting the sample in half, with one sample from 1974Q4 to 1992Q4 and the other from 1993Q1 to 2010Q4, as shown in Tables 3 and 4. Once we correct for benchmark revisions, there is no statistically significant evidence of

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

TABLE 3—BIAS TESTS: 1974Q4 TO 1992Q4

		Surplus		Expenditures		Receipts	
Horizon	Concept	First	Last	First	Last	First	Last
0	Last	0.38	0.80	0.05**	0.03**	0.10*	< 0.01***
	Initial	0.68	0.60	0.12	0.10	0.43	0.05*
	One Year	0.37	0.75	< 0.01***	< 0.01***	0.02**	< 0.01***
	Prebenchmark	0.13	0.23	< 0.01***	< 0.01***	0.12	< 0.01***
1	Last	0.86	0.94	0.61	0.57	0.37	0.32
	Initial	0.95	0.84	0.83	0.79	0.76	0.69
	One Year	0.82	0.89	0.11	0.06*	0.12	0.10
	Prebenchmark	0.62	0.65	0.22	0.16	0.42	0.36
2	Last	0.70	0.76	0.89	0.98	0.32	0.35
	Initial	0.53	0.57	0.64	0.68	0.65	0.71
	One Year	0.66	0.73	0.56	0.40	0.11	0.13
	Prebenchmark	0.88	0.98	0.69	0.56	0.33	0.37
3	Last	0.38	0.45	0.67	0.83	0.16	0.16
	Initial	0.30	0.36	0.64	0.66	0.27	0.28
	One Year	0.36	0.44	0.97	0.76	0.06*	0.06*
	Prebenchmark	0.48	0.59	0.95	0.85	0.18	0.18
4	Last	0.24	0.27	0.61	0.71	0.02**	0.06*
	Initial	0.19	0.21	0.73	0.74	0.06*	0.10
	One Year	0.23	0.26	0.91	0.91	< 0.01***	0.01**
	Prebenchmark	0.31	0.35	0.86	0.96	0.04**	0.07*
1-4	Last	0.63	0.55	0.90	0.92	0.10	0.22
	Initial	0.58	0.48	0.98	0.90	0.24	0.42
	One Year	0.57	0.49	0.66	0.52	0.02**	0.04**
	Prebenchmark	0.76	0.68	0.94	0.92	0.19	0.35

		HEB		HEB6		Unemployment	
Horizon	Concept	First	Last	First	Last	First	Last
0	Last	< 0.01***	< 0.01***	0.01**	0.05**	0.06*	0.05*
	Initial					0.21	0.67
	One Year					0.21	0.67
1	Last	< 0.01***	< 0.01***	< 0.01***	0.02**	0.23	0.09*
	Initial					0.42	0.24
	One Year					0.42	0.24
2	Last	< 0.01***	< 0.01***	< 0.01***	< 0.01***	0.47	0.34
	Initial					0.61	0.50
	One Year					0.61	0.50
3	Last	< 0.01***	< 0.01***	< 0.01***	< 0.01***	0.69	0.57
	Initial					0.81	0.70
	One Year					0.81	0.70
4	Last	< 0.01***	< 0.01***	< 0.01***	< 0.01***	0.64	0.59
	Initial					0.73	0.69
	One Year					0.73	0.69
1-4	Last	< 0.01***	< 0.01***	< 0.01***	< 0.01***	0.39	0.31
	Initial					0.51	0.45
	One Year					0.51	0.45

Note: The figures shown are p-values for tests of the null hypothesis that the mean forecast error is zero. 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 %). Calculations use Newey-West heteroscedasticity and autocorrelation robust standard errors with the number of lags equal to the forecast horizon minus one. The sample period is 1974Q4 to 1992Q4, except for HEB6, for which the sample begins in 1981Q1. *First* and *Last* refer to the first and last FOMC meetings of each quarter.

Table 4—Bias Tests: 1993Q1 to 2010Q4

		Surplus		Expenditures		Receipts	
Horizon	Concept	First	Last	First	Last	First	Last
0	Last	$< 0.01^{***}$	$< 0.01^{***}$	0.51	0.79	$< 0.01^{***}$	$< 0.01^{***}$
	Initial	0.70	0.90	0.06*	0.13	0.31	0.21
	One Year	0.13	0.02**	0.18	0.30	$< 0.01^{***}$	$< 0.01^{***}$
	Prebenchmark	0.47	0.17	0.38	0.15	0.95	0.95
1	Last	0.04**	0.03**	0.67	0.66	0.11	0.12
	Initial	0.87	0.56	0.22	0.17	0.51	0.81
	One Year	0.36	0.41	0.39	0.37	0.08*	0.10
	Prebenchmark	0.63	0.79	0.54	0.47	0.97	0.67
2	Last	0.06*	0.05*	0.85	0.96	0.08*	0.08*
	Initial	0.51	0.54	0.98	0.84	0.30	0.29
	One Year	0.24	0.24	0.90	0.77	0.06*	0.06*
	Prebenchmark	0.32	0.32	0.39	0.42	0.38	0.38
3	Last	0.08*	0.09*	0.73	0.88	0.05*	0.06*
	Initial	0.32	0.39	0.75	0.94	0.15	0.17
	One Year	0.18	0.21	0.95	0.89	0.03**	0.03**
	Prebenchmark	0.24	0.29	0.43	0.51	0.18	0.22
4	Last	0.12	0.12	0.68	0.75	0.05*	0.05*
	Initial	0.30	0.32	0.71	0.80	0.14	0.14
	One Year	0.20	0.20	0.83	0.91	0.03**	0.03**
	Prebenchmark	0.25	0.26	0.47	0.51	0.16	0.16
1-4	Last	0.11	0.09*	0.81	0.85	0.08*	0.07*
	Initial	0.47	0.47	0.81	0.86	0.30	0.29
	One Year	0.23	0.21	0.99	0.94	0.04**	0.04**
	Prebenchmark	0.32	0.31	0.47	0.46	0.28	0.27

		HEB		HEB6		Unemployment	
Horizon	Concept	First	Last	First	Last	First	Last
0	Last	0.43	0.21	0.43	0.21	0.07	0.77
	Initial					0.03**	0.61
	One Year					0.03**	0.60
1	Last	0.49	0.72	0.49	0.72	0.33	0.20
	Initial					0.26	0.13
	One Year					0.26	0.13
2	Last	0.28	0.33	0.28	0.33	0.79	0.62
	Initial					0.73	0.55
	One Year					0.73	0.55
3	Last	0.17	0.23	0.17	0.23	0.99	0.84
	Initial					0.95	0.79
	One Year					0.95	0.79
4	Last	0.15	0.16	0.15	0.16	0.91	0.98
	Initial					0.93	0.98
	One Year					0.93	0.98
1-4	Last	0.29	0.31	0.29	0.31	0.92	0.82
	Initial					0.88	0.77
	One Year					0.88	0.77

Note: The figures shown are p-values for tests of the null hypothesis that the mean forecast error is zero. 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 %). Calculations use Newey-West heteroscedasticity and autocorrelation robust standard errors with the number of lags equal to the forecast horizon minus one. *First* and *Last* refer to the first and last FOMC meetings of each quarter.

bias in the latter part of the sample.[15] In the first part of the sample, however, there is strong evidence of bias in both HEB and HEB6 at all forecast horizons, implying that changing estimates of the natural rate of unemployment were not solely responsible for biased forecasts of the structural surplus, and that this bias continued into the 1980s. In contrast, there is no significant evidence of bias in forecasts of the Surplus or Unemployment. Results for Receipts and Expenditures fall in between these extremes, with strong evidence of bias for nowcasts (0Q horizons) and some additional evidence of bias in forecasts for Receipts due to benchmark revisions.

We now turn to consider some longer-horizon forecasts. As mentioned in Section I, analysis of these data is more difficult as the number and distribution of forecasts over time vary with the forecast horizon. We begin with Table 5, which presents results for the Surplus using all the available data for 4, 6 and 8Q forecasts, as well as results for the period from 1990Q4 onwards.[16] We again see that the choice of outcome measures makes little difference. However, we also see increasingly strong evidence of forecast bias at longer horizons. This impression is potentially misleading, however, as the sample periods are quite different across forecast horizons.

Tables 6 and 7 correct for this in two different ways. Both Tables also provide estimates of the mean forecast bias. Table 6 restricts the sample to forecasts made in the fourth quarter of each year, a time at which Greenbook forecast horizons are typically the longest.[17] As a result we see similar numbers of observations in each column of the table. Table 7 instead imposes the same number of observations in each column.

Both tables continue to show that results are relatively robust to the choice of outcome measure. Furthermore, average forecast errors are positive in every

[15]The p-value in the case of Receipts at the 0-First horizon is 3%. Recalling that we are using tests at six different forecast horizons for five different series for a total of 30 tests, the finding that no p-values are smaller than 3% is consistent with the null hypothesis of no bias in any of the forecasts tested.

[16]Note that 8Q ahead forecasts are only available in the post-1990Q4 period.

[17]This may also control for some differences in the stage of the annual cycle budgetary process across forecasts.

Table 5—Summary Results of Long-Horizon Surplus Bias Tests

		1974Q4–2010Q4		1990Q4–2010Q4	
Horizon	Concept	First	Last	First	Last
4		$N = 111$	$N = 113$	$N = 62$	$N = 62$
	Last	0.06*	0.06*	0.06*	0.05*
	Initial	0.13	0.14	0.17	0.18
	One Year	0.09*	0.10*	0.10*	0.10*
	Prebenchmark	0.14	0.15	0.14	0.14
6		$N = 59$	$N = 74$	$N = 44$	$N = 55$
	Last	0.03**	0.03**	0.04**	0.05**
	Initial	0.04**	0.05*	0.07*	0.09*
	One Year	0.03**	0.04**	0.06*	0.06*
	Prebenchmark	0.04**	0.05*	0.06*	0.08*
8				$N = 13$	$N = 24$
	Last			0.04**	0.03**
	Initial			0.04**	0.03**
	One Year			0.05**	0.03**
	Prebenchmark			0.04**	0.03**

Note: The figures shown are p-values for tests of the null hypothesis that the mean forecast error is zero. 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 %). Calculations use Newey-West heteroscedasticity and autocorrelation robust standard errors with the number of lags equal to the forecast horizon minus one. The sample period is 1974Q4 to 2010Q4, except for HEB6, for which the sample begins in 1981Q1. *First* and *Last* refer to the first and last FOMC meetings of each quarter.

case (implying forecast surpluses exceeded outcomes on average) and increase with forecast horizon, as one might expect. The average error is just over 1% of GDP over the full sample at the 6Q horizon (Table 6), or just over 1.5% in the latter part of the sample at the 8Q horizon (Table 7.) These quantities, while not economically small, are only marginally statistically significant in Table 6 but somewhat more significant in Table 7.

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).[18] Complete results are shown in Appendix Tables 13 to 22 while Table 8 provides a summary. In addition to testing for non-zero medians in the full sample, the latter also provides

[18]These tests control for serial correlation in forecast errors caused by overlapping forecasts and allow for exact inference in small samples. However, due to a lack of observations at the very longest forecast horizons, we only examine forecast horizons up to 6F.

TABLE 6—SUMMARY RESULTS OF LONG-HORIZON SURPLUS BIAS TESTS, USING Q4 DATA ONLY

		1974Q4–2010Q4		1990Q4–2010Q4	
Horizon	Concept	First	Last	First	Last
4		$N = 28$	$N = 28$	$N = 16$	$N = 16$
	Last	0.007	0.005	0.008	0.008
		[0.05*]	[0.08*]	[0.12]	[0.08*]
	Initial	0.006	0.005	0.006	0.006
		[0.07*]	[0.10]	[0.21]	[0.16]
	One Year	0.007	0.005	0.008	0.008
		[0.06*]	[0.08*]	[0.14]	[0.10]
	Prebenchmark	0.006	0.004	0.007	0.006
		[0.10*]	[0.14]	[0.20]	[0.16]
6		$N = 23$	$N = 25$	$N = 16$	$N = 16$
	Last	0.009	0.009	0.012	0.012
		[0.08*]	[0.06*]	[0.09*]	[0.09*]
	Initial	0.009	0.008	0.011	0.010
		[0.09*]	[0.07*]	[0.12]	[0.12]
	One Year	0.009	0.009	0.012	0.012
		[0.08*]	[0.07*]	[0.10*]	[0.10*]
	Prebenchmark	0.008	0.008	0.011	0.010
		[0.11]	[0.08*]	[0.12]	[0.12]
8				$N = 11$	$N = 12$
	Last			0.016	0.015
				[0.12]	[0.06*]
	Initial			0.016	0.015
				[0.13]	[0.07*]
	One Year			0.015	0.015
				[0.14]	[0.07*]
	Prebenchmark			0.015	0.014
				[0.13]	[0.07*]

Note: The figures shown are mean forecast errors (forecast - actual) with p-values in brackets below for tests of the null hypothesis that the mean forecast error is zero. 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 %). Calculations use Newey-West heteroscedasticity and autocorrelation robust standard errors with the number of lags equal to the forecast horizon minus one. The sample period is 1974Q4 to 2010Q4, except for HEB6, for which the sample begins in 1981Q1. *First* and *Last* refer to the first and last FOMC meetings of each quarter.

results for the period 1986Q1–2006Q4 corresponding to the Great Moderation. Results presented in the Appendix examine the sensitivity these results to the test statistic used, the sample period and the outcome measured to construct the forecast errors. Except as noted below, however, results tended to be robust to these choices.

Table 8 shows that there is little evidence of median bias among the three deficit measures HEB, HEB6, and SURPLUS except in the case of short-horizon forecasts of HEB where overly optimistic forecasts were relatively more common.

TABLE 7—SUMMARY RESULTS OF LONG-HORIZON SURPLUS BIAS TESTS, EQUAL N IN EACH COLUMN

		1974Q4–2010Q4		1990Q4–2010Q4	
Horizon	Concept	First	Last	First	Last
4		$N = 59$	$N = 73$	$N = 13$	$N = 24$
	Last	0.008	0.008	0.011	0.011
		[0.04**]	[0.03**]	[0.04**]	[0.02**]
	Initial	0.006	0.006	0.010	0.009
		[0.11]	[0.09*]	[0.06*]	[0.05**]
	One Year	0.007	0.007	0.011	0.011
		[0.06*]	[0.05*]	[0.03**]	[0.03**]
	Prebenchmark	0.006	0.006	0.010	0.010
		[0.10*]	[0.08*]	[0.06*]	[0.04**]
6		$N = 59$	$N = 74$	$N = 13$	$N = 24$
	Last	0.012	0.011	0.018	0.015
		[0.03**]	[0.03**]	[0.03**]	[0.05**]
	Initial	0.011	0.010	0.016	0.013
		[0.04**]	[0.05*]	[0.06*]	[0.07*]
	One Year	0.012	0.011	0.018	0.015
		[0.03**]	[0.04**]	[0.03**]	[0.05**]
	Prebenchmark	0.011	0.010	0.017	0.013
		[0.04**]	[0.05*]	[0.06*]	[0.08*]
8				$N = 13$	$N = 24$
	Last			0.018	0.019
				[0.04**]	[0.03**]
	Initial			0.018	0.018
				[0.04**]	[0.03**]
	One Year			0.018	0.019
				[0.005**]	[0.03**]
	Prebenchmark			0.017	0.018
				[0.04**]	[0.03**]

Note: The figures shown are forecast errors (forecast - actual) with p-values in brackets below for tests of the null hypothesis that the mean forecast error is zero. 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 %). Calculations use Newey-West heteroscedasticity and autocorrelation robust standard errors with the number of lags equal to the forecast horizon minus one. The sample period is 1974Q4 to 2010Q4, except for HEB6, for which the sample begins in 1981Q1. *First* and *Last* refer to the first and last FOMC meetings of each quarter.

Even there, however, such behavior appeared to be limited to the pre-1986 period, during which the benchmark rate of unemployment was gradually increased over time. There was also evidence of median bias in the opposite direction (overly pessimistic forecasts were too common) at the very longest forecast horizon (6F) across all three variables. Again, there was little statistically significant evidence of such behavior in the post-1985 period, although this appears to be due to a reduction in the test's power in a shorter sample. As shown in Appendix Tables 13 to 16, the use of alternative test statistics produced more significant evidence

Table 8—Forecast Error Sign Tests

Series:	HEB		HEB6		Surplus	
Sample:	Full	86-'06	Full	86-'06	Full	86-'06
h=0L	0.655***	0.518	0.49	0.518	0.462	0.482
h=0F	0.600**	0.494	0.462	0.494	0.434	0.447
h=1L	0.590**	0.417	0.438	0.417	0.424*	0.417
h=1F	0.604***	0.464	0.472	0.464	0.417*	0.417
h=2L	0.648**	0.561	0.380	0.561	0.380	0.366
h=2F	0.620*	0.439	0.423	0.439	0.366*	0.39
h=3L	0.617	0.407	0.404	0.407	0.383	0.370
h=3F	0.638	0.556	0.404	0.556	0.362	0.370
h=4L	0.571	0.400	0.371	0.400	0.314	0.250
h=4F	0.543	0.500	0.371	0.500	0.343	0.300
h=5L	0.571	0.563	0.357	0.563	0.321	0.313
h=5F	0.393	0.313	0.321	0.313	0.286	0.250
h=6L	0.348	0.308	0.261	0.308	0.217*	0.154
h=6F	0.174**	0.154	0.174**	0.154	0.174**	0.154
Series:	Expenditures		Receipts		Unemployment	
Sample:	Full	86-'06	Full	86-'06	Full	86-'06
h=0L	0.579**	0.553	0.607***	0.647***	0.552	0.565
h=0F	0.600**	0.588*	0.531	0.553	0.648***	0.682***
h=1L	0.514	0.512	0.500	0.464	0.611***	0.667***
h=1F	0.535	0.548	0.507	0.476	0.681***	0.762***
h=2L	0.451	0.463	0.408	0.341	0.690***	0.732***
h=2F	0.408	0.537	0.423	0.415	0.704***	0.780***
h=3L	0.404	0.407	0.362	0.333	0.723***	0.815***
h=3F	0.383	0.407	0.447	0.370	0.702***	0.778***
h=4L	0.200***	0.250	0.343	0.400	0.714**	0.750**
h=4F	0.229***	0.300	0.314	0.550	0.686*	0.750**
h=5L	0.214**	0.250	0.357	0.313	0.643	0.750
h=5F	0.179***	0.250	0.321	0.313	0.643	0.750
h=6L	0.087***	0.154	0.304	0.385	0.391	0.692
h=6F	0.043***	0.077**	0.217*	0.231	0.217*	0.615

Note: The figures in this table are fraction of forecast errors (Forecast - Outcome) greater than zero. Forecast horizons (h) range from 0 quarters (nowcasts) to 6 and from the first (F) and last (L) FOMC meetings of the quarter. ***/**/* indicates rejection of the null hypothesis at the 1/5/10% significance level using the Sign Test statistic. At forecast horizons of 1L or more, figures incorporate the Bonferroni correction for overlapping observations proposed by Campbell and Dufour (1991) and Campbell and Ghysels (1995).

Outcomes for most variables are measured using the last value reported prior to a benchmark revision. Outcomes for HEB and HEB6 are the last values reported in the Greenbook. Outcomes for the unemployment rate are Current Vintage.

The full sample starts in 1974Q4 and ends in 2010Q4. The '86–'06 sample covers the 1986Q1 to 2006Q4 period (inclusive) and captures the Great Moderation.

of median bias, but such evidence again disappeared after omitting the pre-1986 part of the sample.

Table 8 also shows little evidence of median bias in Receipts except at the very shortest horizon nowcasts, where again overly optimistic forecasts were relatively more common. Nowcasts for Expenditures showed a similar median bias, while longer horizon Expenditure forecasts were commonly too low in the pre-1986 period.

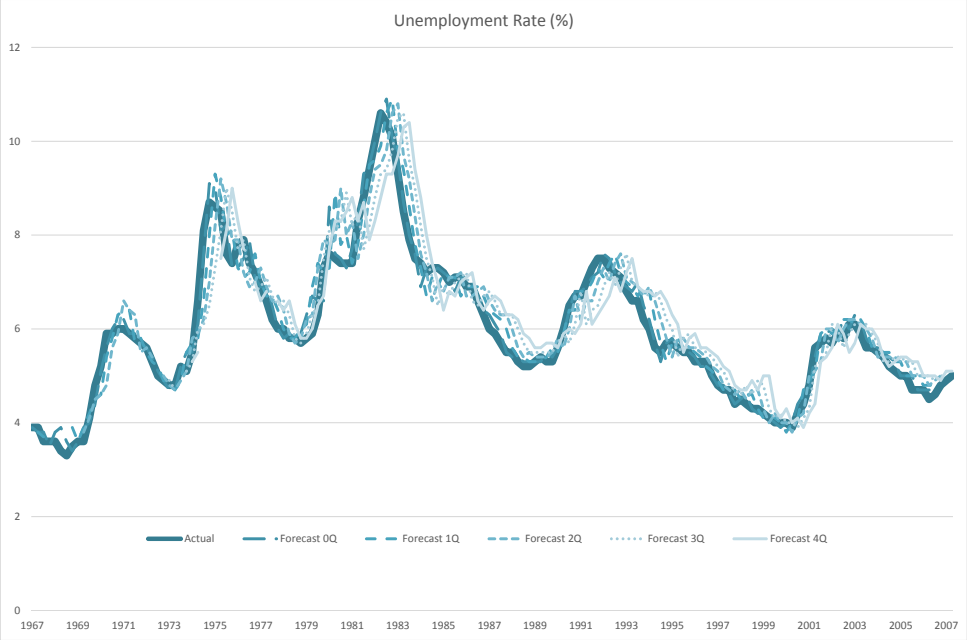

Figure 5. Unemployment Rate Forecasts and Outcomes

Far and away the most evidence of median bias was found in all but the longest horizon unemployment rate forecasts, which typically were too pessimistic (forecast unemployment rates were too high too frequently) both in the full sample and during the Great Moderation. However, given the lack of any significant evidence

of bias in unemployment rate forecasts in Table 2 for horizons longer than 1Q, this is most likely due to the relatively skewed distribution of unemployment rate changes rather than true forecast bias. Figure 5 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.

A. 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 that other forecasts are systematically over-optimistic in advance of elections, we might expect the Greenbook forecasts to be an exception as they are not publicly 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 ($POST-ELECTION$).[19] For simplicity, we test only forecast errors using our "best" measure of forecast outcomes; Pre-benchmark estimates for Expenditures, Receipts, and Surplus, Final for the unemployment rate, and the Last Greenbook value for HEB and HEB6. To allow for sufficient degrees of freedom, we consider only forecast horizons from zero to four quarters ahead and test the period 1974Q4–2006Q4.

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

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.[20]

B. Turning Points and Structural Surpluses

Forecasters and policymakers are often particularly concerned about the ability of their forecasts to capture business cycle turning points. One reason for this may be that they feel errors are particularly costly at such times. However, we would expect that even an efficient forecast will appear to be biased around turning points. The reason for this is that turning points are identified only with a (sometimes substantial) delay. This means that they are not part of the information that was available to forecasters. If we pick turning points *ex post*, we should expect forecasts made around peaks to be overly optimistic on average and those around troughs to be similarly too pessimistic on average.

Forecasts of the structural surplus (HEB) are a possible exception to this rule. In particular, HEB tries to capture the stance of fiscal policy by purging the budget surplus of the "direct effect" of cyclical shocks. We should therefore expect to find turning-point bias in HEB forecasts only to the extent that *discretionary* fiscal policy responds to such shocks. The extent to which we observe such bias in practice will depend on (a) whether HEB adequately estimates and compensates for the direct effects of cyclical shocks, (b) the time it takes to recognize that a turning point has passed, and (c) the time it takes discretionary fiscal policy to

[20]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.

react to a turning point.

We are not aware of previous empirical studies that have examined how business cycle turning points affect estimated and forecast structural balances. This may simply reflect the fact that most previous studies either did not include estimates of structural balances or covered too few business cycles to make a meaningful comparison. We therefore investigated the behavior of HEB forecast errors around NBER business cycle turning points.

Our sample covers six business cycle peaks (November 1973, January 1980, July 1981, July 1990, March 2001, and December 2007) and seven troughs (November 1970, March 1975, July 1980, November 1982, March 1991, November 2001, and June 2009). For each date, we took the forecasts from the last FOMC meeting prior to the end of month containing the turning point. Because of the small sample size, we made no formal attempt to test for bias. Our results are summarized in Figures 6 and 7.

The colored narrow lines show the forecast error by forecast horizon for each business cycle, while the thicker black line shows their average across the business cycles. Surprisingly, the two figures present similar results. Although individual cycles are widely dispersed about the sample average, both peaks and troughs show average forecast errors that are typically quite small (<0.5% of GDP) at the shortest horizons but increase fairly steadily, exceeding 2.0% of GDP about a year after cycle peaks and 1.0% after cycle troughs. These positive errors imply that the forecasts of structural surpluses were overly optimistic. This result seems distinct from any full-sample bias in HEB forecasts (which is typically much less than half the size but is of the same sign).[21]

Instead, HEB estimates seem to be sharply revised downwards (i.e., toward larger *structural* deficits) in the immediate aftermath of business cycle peaks and troughs. This pattern is more clearly visible in Figure 8. Figure 9 provides a

[21]The fact that both peaks and troughs produce mean forecast errors of the same sign implies that the ex-post identification of turning points cannot be responsible for this result.

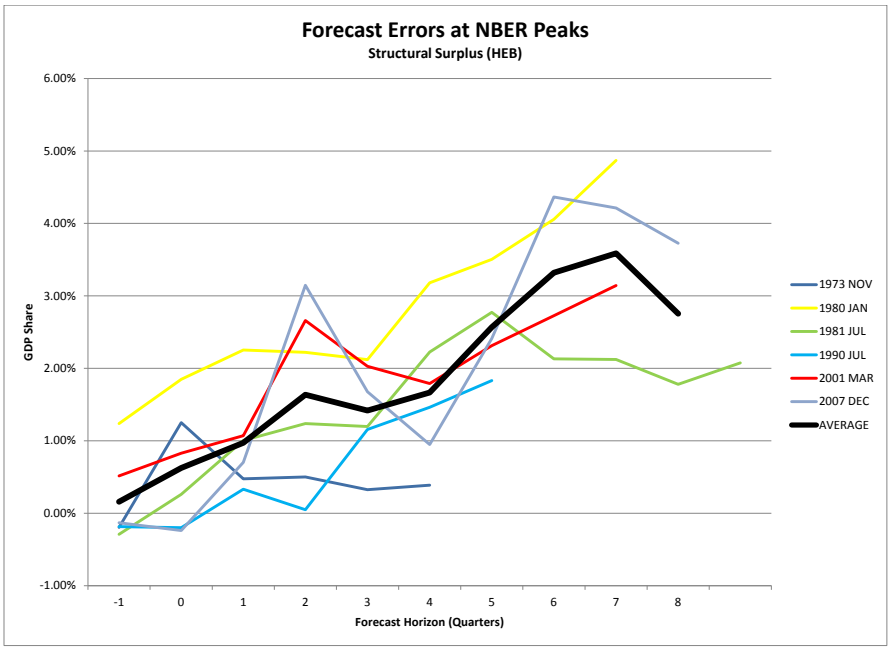

FIGURE 6. HEB—FORECAST ERRORS AT BUSINESS CYCLE PEAKS

scatterplot of unemployment rate forecast errors against those of HEB. While positive (i.e., overly pessimistic) unemployment rate forecast errors show no clear association with either positive or negative HEB forecast errors, negative unemployment rate forecast errors (i.e., overly optimistic forecasts) are associated with positive (i.e., overly optimistic) HEB forecasts. This relationship is particularly strong for larger unemployment rate surprises, such as we might expect as the economy enters a recession.

Taken at face value, the positive mean forecast errors at business cycle peaks and troughs appear to be evidence that

- discretionary fiscal policy consistently becomes more counter-cyclical during recessions

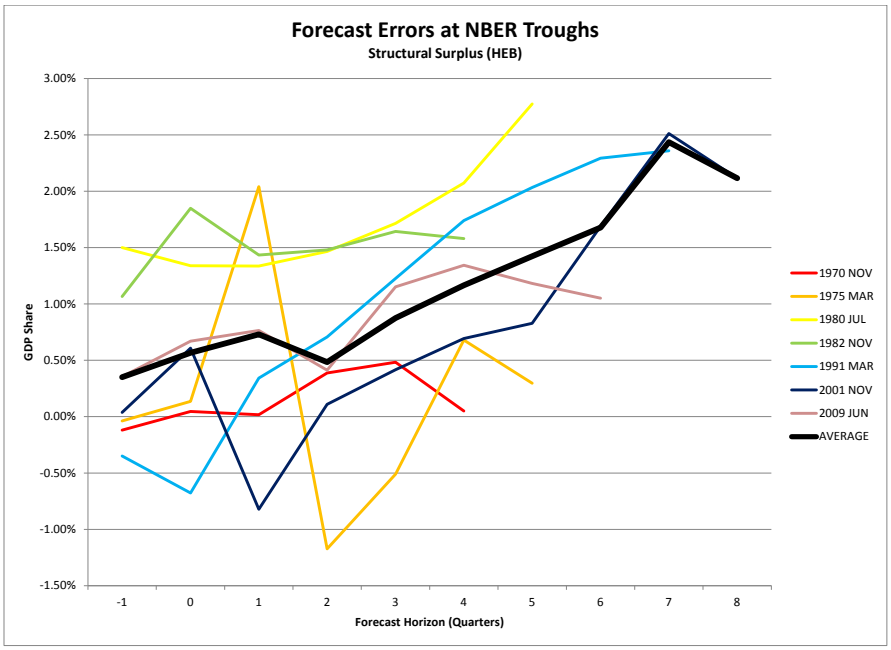

FIGURE 7. HEB—FORECAST ERRORS AT BUSINESS CYCLE TROUGHS

- it continues to become more stimulative (and therefore pro-cyclical) during the early stages of a recovery
- the former effect seems somewhat stronger than the latter.

This is consistent with the traditional view that U.S. fiscal policy since the mid-20th century has been largely counter-cyclical, rather than the more modern view that it has instead tended to be acyclical or pro-cyclical. However, other interpretations are also possible. For example, given the lags between business cycle peaks and the recognition that a recession has begun, it is hard to credit the positive forecast errors at horizons of less than 3 or 4 quarters as being due to a deliberate and discretionary fiscal policy. It might instead be possible that in the aftermath of recessions, the Board has tended to systematically re-estimate

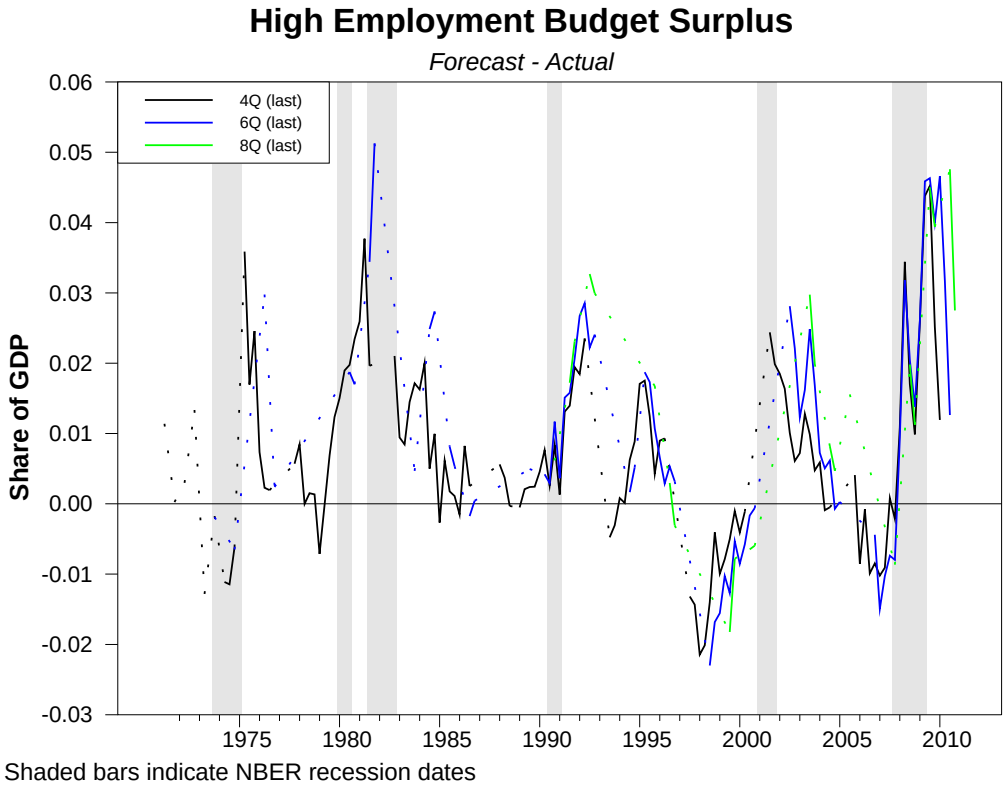

FIGURE 8. HEB—FORECAST ERRORS (WITH NBER BUSINESS CYCLE DATES)

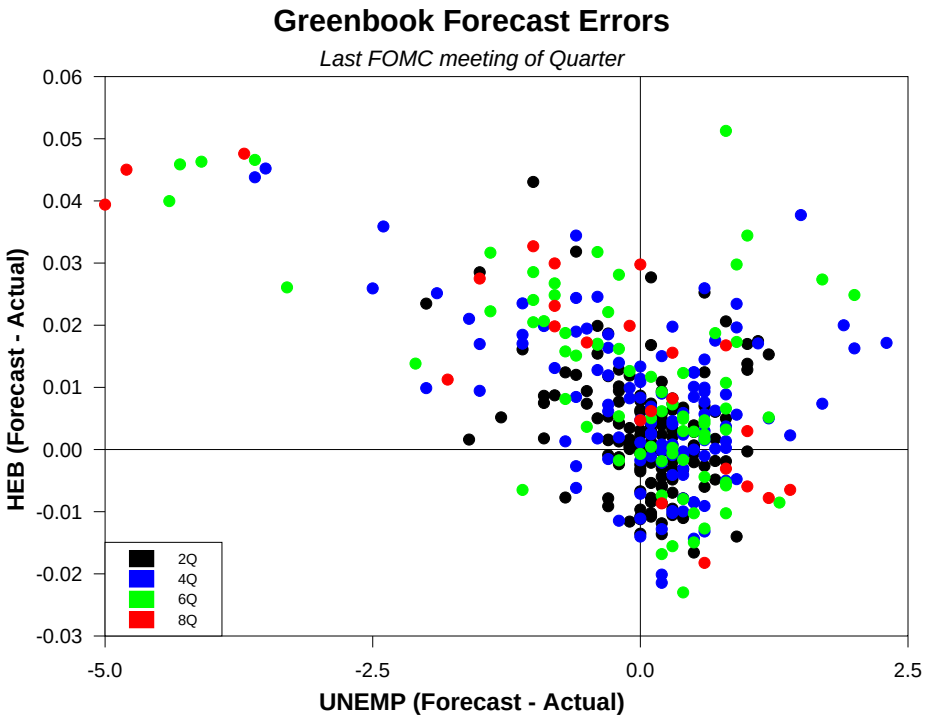

FIGURE 9. HEB—FORECAST ERRORS AT BUSINESS CYCLE TROUGHS

the relationship between unemployment rates and the actual surplus.

IV. 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 on the one hand and data that were in the information set when each Greenbook forecast was produced on the other. Because of the timing requirements, it is crucial that real-time data be used in such an exercise.

A. Forecast Error Persistence

Starting with Scotese (1995), multiple studies have found that Greenbook forecast errors tend to be serially correlated.[22] 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 showed less or no evidence of inefficiency.[23] 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.[24]

[22]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.

[23]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.

[24]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. However, this may be due to the asymmetric distribution of unemployment rate changes noted above rather than forecast inefficiency *per se*.

B. The Fed Funds Rate

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 Survey of Professional Forecasters (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 9—SUMMARY RESULTS OF EFFICIENCY TESTS

		Surplus		Expenditures		Receipts	
Horizon	Concept	First	Last	First	Last	First	Last
0	Last	0.21	0.14	0.55	0.85	0.09	0.06
	Initial	0.21	0.09	0.54	0.91	0.13	0.08
	One Year	0.60	0.56	0.70	0.96	0.49	0.49
	Prebenchmark	0.26	0.16	0.75	0.87	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	0.33	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	0.11	0.16
	Initial	0.66	0.63	0.09	0.10	0.09	0.11
	One Year	0.53	0.51	0.16	0.20	0.33	0.42
	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	0.34	0.07	0.45	0.92	0.86	0.37
	Initial					0.99	0.20
	One Year					0.99	0.20
2	Last	0.82	0.46	0.26	0.57	0.28	0.24
	Initial					0.24	0.28
	One Year					0.24	0.28
4	Last	0.71	0.42	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.

Table 9 shows the results of the efficiency tests. 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.[25] Thus, the Ball and Croushore (2003) results on the inefficiency of the SPF forecasts do not carry over to fiscal forecasts in the Greenbook.

C. 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 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, many 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 with the corresponding Greenbook forecast by averaging the four quarterly Greenbook forecasts to compute the implied annual forecast.[26] Both sets of forecasts are compared in Table 10. 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 onward. We also compare these Greenbook and CBO forecasts with those of a

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

[26] 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.

4Q and 8Q random walk (RW) forecast.[27]

TABLE 10—GREENBOOK VERSUS CBO AND RANDOM WALK

Variable	**Surplus**		**Receipts**		**Expenditures**	
Horizon (Years)	**0**	**1**	**0**	**1**	**0**	**1**
RMSFE - Greenbook	0.0086	0.0141	0.0049	0.0103	0.0052	0.0088
RMSFE - CBO	0.0092	0.0171	0.0067	0.0121	0.0058	0.0107
RMSFE - Random Walk	0.0110	0.0178	0.0068	0.0128	0.0064	0.0091
Greenbook versus CBO						
H_0 : Equal Quadratic Loss	0.726	0.251	**0.031**	**0.034**	0.342	0.142
H_0 : Equal Absolute Loss	0.578	0.221	**0.020**	0.156	0.671	0.333
H_0 : GB encompasses CBO	0.465	0.378	0.800	0.099	0.564	0.375
H_0 : CBO encompasses GB	0.252	0.185	**0.003**	**0.015**	**0.017**	0.071
Greenbook versus Random Walk						
H_0 : Equal Quadratic Loss	0.337	0.136	0.124	0.121	0.140	0.872
H_0 : Equal Absolute Loss	0.203	0.163	0.073	0.096	0.189	0.851
H_0 : GB encompasses RW	0.328	0.552	0.900	0.211	0.514	0.552
H_0 : RW encompasses GB	0.076	0.052	**0.042**	0.079	**0.026**	0.139

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 Harvey, Leybourne and Newbold (1997) modified Diebold-Mariano test statistic of the corresponding null hypothesis.
Figures in the final two rows are p-values for tests of the null hypothesis of forecast encompassing using the statistic proposed by Harvey, Leybourne and Newbold (1998) and incorporate their proposed small-sample adjustment.
Boldface denotes p-values $< 5\%$.

Table 10 compares the performance of the Greenbook, the CBO, and the RW forecasts in a number of ways.[28] The first three lines simply report the root-mean-squared forecast errors. We see that Greenbook forecasts are always the most accurate of the three.

The middle section of the table tests the relative performance of the Greenbook and CBO forecasts. The first line tests the null hypothesis that the two forecasts have equal mean-squared forecast errors and reports the associated p-values.[29] We find that the Greenbook forecasts are significantly more accurate only for forecasts of receipts; differences in the accuracy of their forecasts of the surplus

[27]The random walk forecast holds constant the value of forecasted variable *as a fraction of GDP*. It is calculated using the pre-benchmark vintage of the outcome series and so ignores the possible effects of data revision. For comparability, RW forecasts are calculated only for those periods in which the two other forecasts are available.

[28]In interpreting these results, it should be recalled that these forecasts condition on distinctly different assumptions, as discussed above.

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

or expenditures were not significant. The following line repeats the test, replacing the squared forecast error with mean absolute error. Now only the current-year forecasts of receipts show a significant difference. The two final lines test whether one forecast forecast-encompasses the other.[30] We find no evidence that the CBO forecast contains information missing from the Greenbook forecast, while the Greenbook forecasts for Receipts and Expenditures (but not the Surplus) contain significant evidence missing from the CBO forecasts.[31]

The final section of the table tests the relative performance of the Greenbook and the random walk forecasts using the same types of tests used above. There is now no statistically significant difference in mean-squared or mean-absolute forecast error (although in the latter case, evidence for receipts is significant at the 10% level.) There is no evidence that the random walk forecasts contain information lacking from the Greenbook forecasts, while we find evidence of the reverse that is significant at the 10% level in 5 of the 6 cases tested.

V. Summary and Conclusions

Our examination of the Greenbook's fiscal forecasts should help us better understand the size and characteristics of the unanticipated fiscal policy shocks confronting monetary policymakers. In doing so, we find that these shocks are often large (surpassing 1–2% of GDP) and skewed, with relatively larger downside risks (i.e. larger deficits than expected). Furthermore, while the overall predictability of the unemployment rate improved after 1990, the predictability of our fiscal variables deteriorated, often substantially. Despite this, the Board staff's forecasts gave lower root-mean-squared forecast errors than either those of the CBO or the assumption of "no change" (i.e., a random walk). Furthermore, the latter two sets of forecasts never appeared to contain information that would

[30] 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.

[31] 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.

significantly improve those in the Greenbook. There was also limited evidence of forecast bias, and the majority of that evidence was confined to the pre-1993 period.

Taken together, these results suggest that forecasts provided to the FOMC are relatively efficient, and that their lack of accuracy in recent decades reflects a fundamental increase in the size of fiscal shocks and the volatility of discretionary fiscal policy. We also find that cyclically adjusted deficit forecasts appear to be over-optimistic around both business cycle peaks and troughs. This appears to imply that U.S. discretionary fiscal policy has been consistently counter-cyclical during economic downturns and somewhat pro-cyclical (i.e., stimulative) during the early stages of economic recoveries.

REFERENCES

Aruoba, S. Boragan. 2008. "Data Revisions Are Not Well Behaved." *Journal of Money, Credit, and Banking*, 40: 319–340.

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.

Bernanke, Ben S. 2017. "The Fed and Fiscal Policy." *Brookings Blog.*

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 Non-parametric Approach to the Mankiw-Shapiro Problem." *Economics Letters*, 35(3): 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 Simon van Norden. 2017. "Fiscal Forecasts at the FOMC: Evidence from the Greenbooks." *Review of Economics and Statistics*, forthcoming.

Harvey, David S., Stephen J. Leybourne, and Paul Newbold. 1997. "Testing the Equality of Prediction Mean Squared Errors." *International Journal of Forecasting*, 13: 281–291.

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

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.

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.

VI. Appendix

A. Benchmark Revisions

We use the extent of revision to define those that we treat as *benchmark* revisions. Benchmark revisions are those that affect a significant portion of the published history of a time series. For example, quarterly U.S. National Accounts are available starting from 1946Q1. Revisions that do not affect the published estimates for more than five years are therefore not considered benchmark revisions. Changes in seasonal adjustment factors, although they may occur many years after the fact, are not counted as benchmark revisions. Changes in base years (for real values), or the change from fixed-weight to chain-weighted values, or the change from GNP to GDP, are all examples of benchmark changes. This definition of benchmark revision has at least two important advantages.

1. It is a simple, transparent and objective way to determine which revisions are treated as benchmark revisions.

2. It implicitly relies on the judgment of the statistical agency to determine which methodological or conceptual changes are important enough to be considered benchmark changes. In effect, if the statistical agency judges that historical estimates are sufficiently comparable to current estimates that no revision to the former is required, no benchmark revision has occurred.

This definition also has at least one important drawback: since no official series is published for HEB, no long time series are available to identify benchmark changes. As we describe below, we therefore treat HEB estimates somewhat differently.

The economic importance of benchmark revisions varied vastly across our series, as Croushore and van Norden (2017) describe in greater detail. At one extreme, benchmark revisions in the unemployment rate were rare and trivial. In contrast, the redefinition of the government accounts had an important impact on our fiscal variables. We discuss the economic importance of benchmark revisions in

the next subsection. Table 11, taken from the appendix to Croushore and van Norden (2017), shows the dates at which benchmark revisions were first published for each series.

TABLE 11—PRE-BENCHMARK-REVISION DATES FOR QUARTERLY NATIONAL ACCOUNTS

Last Quarter	Last ALFRED Vintage	Last FOMC Date
1975:3	Dec. 19, 1975	Dec. 10, 1975
1980:3	Nov. 19, 1980	Dec, 12, 1980
1985:3	Nov. 20, 1985	Dec. 11, 1985
1991:2	Aug. 28, 1991	Oct. 30, 1991
1995:2	Oct. 27, 1995	Dec. 14, 1995
1999:2	Sep. 30, 1999	Sep. 29, 1999
2003:3	Nov. 25, 2003	Dec. 03, 2003
2009:1	Jun. 25, 2009	Jun. 17, 2009

Note: This table gives the dates of publication for the last estimates prior to benchmark revisions of the National Accounts. The first column gives the last time period to which those estimates correspond. The second column gives the date at which those estimates were published. The last column gives the date of the last FOMC meeting prior to the publication of the benchmark revision. These dates apply to figures from the Quarterly National Accounts as based on original data vintages from ALFRED and the FRB Philadelphia Real-Time Data Set for Macroeconomists. The 1995 benchmark revision of Expenditures occurred slightly after the revision of the other series; its last pre-benchmark-revision quarter was 1995:3 which was published on October 27, 1995. The last FOMC meeting using this estimate was that of December 1995.

Values forecast prior to benchmark revision are not comparable to outcomes measured after a benchmark revision. For that reason, whenever a forecast or nowcast is made for an outcome that will only be observed after a benchmark revision has occurred, we drop those forecast errors from our data set. For example, the Greenbook for the FOMC meeting in October 1975 contained nowcasts and forecasts for the period 1975Q4-1976Q4. Estimates for most of these outcomes were only published after the benchmark revision which was first released on January 20, 1976. Therefore, for the series affected by those benchmark changes, those forecast errors were replaced by a missing value code.

B. Forecast Accuracy

Table 12 provides details on forecast error variances.

Table 12—Forecast Error Variances

Horizon	Expenditures	Receipts	Surplus	HEB	HEB6	Unemployment
1967Q3–1990Q4						
0L	0.038	0.059	0.036	0.128	0.086	0.006
0F	0.073	0.106	0.078	0.166	0.089	0.027
1L	0.105	0.136	0.122	0.219	0.093	0.045
1F	0.141	0.198	0.177	0.230	0.085	0.080
2L	0.159	0.269	0.221	0.264	0.071	0.116
2F	0.192	0.291	0.246	0.276	0.068	0.151
3L	0.227	0.286	0.278	0.308	0.087	0.194
3F	0.275	0.292	0.303	0.310	0.088	0.236
4L	0.274	0.361	0.306	0.335	0.056	0.271
4F	0.340	0.330	0.332	0.406	0.084	0.299
5L	0.308	0.219	0.180	0.411	0.090	0.299
5F	0.323	0.125	0.141	0.498	0.229	0.330
6L	0.287	0.125	0.122	0.687	0.303	0.336
6F	0.342	0.099	0.164	0.868	0.326	0.377
1991Q1–2006Q4						
0L	0.049	0.098	0.040	0.110	0.149	0.003
0F	0.059	0.107	0.053	0.128	0.174	0.009
1L	0.066	0.137	0.086	0.170	0.230	0.012
1F	0.069	0.174	0.096	0.160	0.217	0.024
2L	0.100	0.307	0.188	0.240	0.325	0.029
2F	0.117	0.321	0.212	0.232	0.314	0.042
3L	0.168	0.465	0.325	0.308	0.417	0.051
3F	0.194	0.481	0.357	0.304	0.410	0.078
4L	0.258	0.635	0.497	0.403	0.544	0.091
4F	0.302	0.644	0.545	0.416	0.563	0.122
5L	0.357	0.768	0.662	0.497	0.671	0.128
5F	0.401	0.802	0.726	0.519	0.701	0.158
6L	0.430	0.783	0.745	0.516	0.697	0.165
6F	0.477	0.832	0.808	0.599	0.809	0.196
7L	0.523	0.573	0.723	0.461	0.624	0.195
7F	0.564	0.470	0.696	0.567	0.767	0.222
8L	0.558	0.505	0.856	0.583	0.789	0.231
8F	0.677	0.351	0.836	0.477	0.645	0.280

Note: Forecast error variances are shown as a fraction of the unconditional variance of the underlying series over the period 1967Q3–2010Q4. Because HEB and HEB6 are identical in the post-1990 period, their figures above differ only to the extent that their variances differ in the 1967–2010 sample.
Forecasts are taken from the first FOMC meeting in 1967Q3 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.
Forecasts with horizons longer than 6Q were not available for all series prior to 1991.

C. Distribution of Forecast Errors

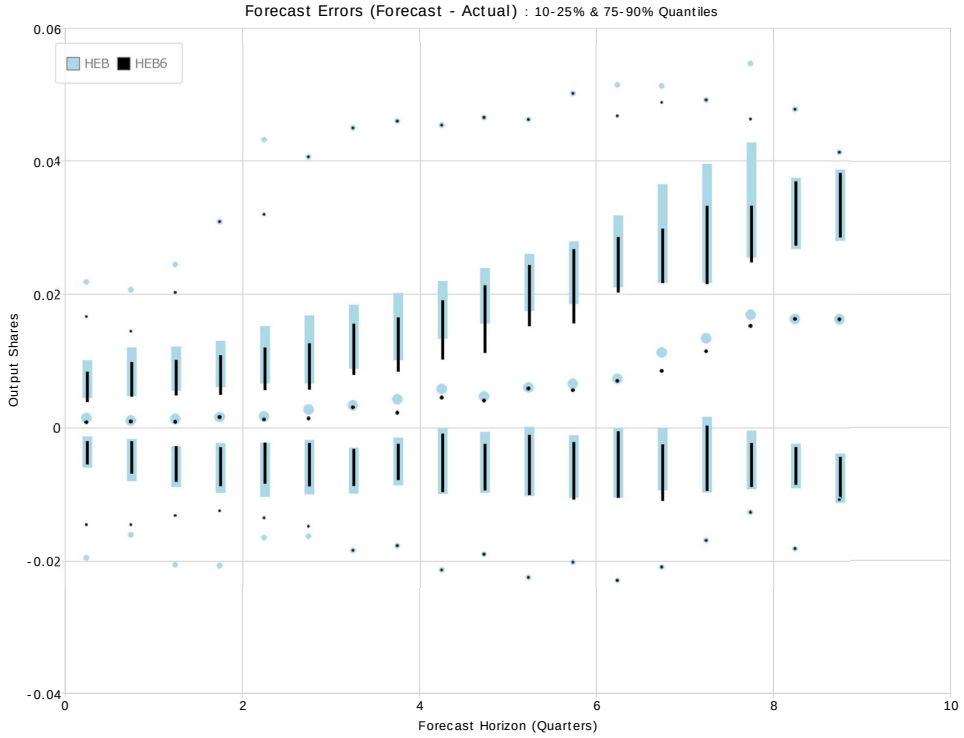

FIGURE 10. FORECAST ERRORS FOR HEB6 AND HEB

Note: The simplified box plots above compare forecast error quantiles for HEB6 and HEB. At each forecast horizon, vertical lines link the 90th to the 75th percentiles as well as the 25th to the 10th percentile. Dots indicate the median, maximum, and minimum forecast errors.

D. Tests for Median Bias

Notes for Tables 13 to 22:

- Sample periods start in 1974Q4 and end in 2010Q4 unless otherwise indicated.
- Forecast horizons (h) range from 0 quarters (nowcasts) to 4 and from the first (F) and last (L) FOMC meetings of the quarter.
- All results are for tests of the null hypothesis that the median of the forecast errors is equal to zero.
- ***/**/* indicates rejection of the null hypothesis at the 1/5/10% significance level.
- The **Sign** columns show the Sign Test statistics. Values greater/less than 0.5 indicate that a greater/lesser proportion of forecast errors were positive.
- The **Signed-Rank** columns show the Wilcoxon Signed-Rank statistics. Larger values are associated with distributions skewed toward positive values.
- Outcome measures used to calculate forecast errors include the last Greenbook reported value (Last), the initial estimate (First), the estimate as of one year after release after the initial estimate (1 Yr.), the last estimate prior to Benchmark revision (Pre-B.), and the current vintage (CV).
- At forecast horizons greater than one quarter, figures incorporate the Bonferroni correction for overlapping observations proposed by Campbell and Dufour (1991) and Campbell and Ghysels (1995).
- Note that the test has relatively little power to reject H_0 for the longest horizon forecasts in shorter samples (particularly the pre-1986 sample.)

TABLE 13—TESTS OF MEDIAN FORECAST BIAS: HEB, HEB6

Forecast Horizon	Full Sample		Pre-1991Q1		Post-1990Q4	
	Sign	Signed-Rank	Sign	Signed-Rank	Sign	Signed-Rank
HEB (Last)						
h=0L	0.655***	7570.0***	0.754***	1790.0***	0.575	1973.0*
h=0F	0.600**	7150.0***	0.692***	1686.0***	0.525	1904.0
h=1L	0.590**	6354.0**	0.703***	1579.0***	0.494	1562.0
h=1F	0.604***	6468.0**	0.750***	1616.0***	0.494	1601.0
h=2L	0.648**	1505.0	0.774***	378.0**	0.513	339.0
h=2F	0.620*	1517.0	0.774***	357.0*	0.436	338.0
h=3L	0.617	619.0	0.750**	141.0	0.400	123.0
h=3F	0.638	638.0	0.800***	143.0	0.400	128.0
h=4L	0.571	298.0	0.733	77.0	0.526	70.0
h=4F	0.543	274.0	0.733	77.0	0.368	68.0
h=5L	0.571	128.0	0.667	15.0	0.333	30.0
h=5F	0.393	109.0	0.250	9.0*	0.267	26.0
h=6L	0.348	45.0**	0.111	1.0*	0.333	21.0
h=6F	0.174**	18.0***	0.111	1.0*	0.167	7.0*
HEB6 (Last)						
h=0L	0.490	4549.0	0.385*	545.0***	0.575	1973.0*
h=0F	0.462	4364.0*	0.385*	538.0***	0.525	1904.0
h=1L	0.438	3809.0***	0.359**	508.0***	0.494	1562.0
h=1F	0.472	3959.0**	0.453	572.0***	0.494	1601.0
h=2L	0.380	783.0***	0.258**	75.0***	0.513	339.0
h=2F	0.423	819.0**	0.419	118.0**	0.436	338.0
h=3L	0.404	330.0**	0.350	40.0**	0.400	123.0
h=3F	0.404	330.0**	0.350	38.0**	0.400	128.0
h=4L	0.371	147.0**	0.200	6.0***	0.526	70.0
h=4F	0.371	143.0**	0.200	6.0***	0.368	68.0
h=5L	0.357	82.0**	0.167	3.0**	0.333	30.0
h=5F	0.321	71.0**	0.083**	1.0**	0.267	26.0
h=6L	0.261	38.0**	0.111	1.0*	0.333	21.0
h=6F	0.174**	18.0***	0.111	1.0*	0.167	7.0*

TABLE 14—TESTS OF MEDIAN FORECAST BIAS: HEB, HEB6

Forecast Horizon	Full Sample to 2006		Pre-1986Q1		1986Q1–2006Q4	
	Sign	Signed-Rank	Sign	Signed-Rank	Sign	Signed-Rank
HEB (Last)						
h=0L	0.636***	5782.0***	0.864***	918.0***	0.518	1964.0
h=0F	0.589**	5469.0***	0.773***	832.0***	0.494	1956.0
h=1L	0.563	4763.0	0.837***	784.0***	0.417	1476.0
h=1F	0.586**	4902.0*	0.814***	769.0***	0.464	1612.0
h=2L	0.651**	1182.0	0.810***	187.0**	0.561	320.0
h=2F	0.619*	1183.0	0.810***	173.0*	0.439	342.0
h=3L	0.595	488.0	0.846**	66.0	0.407	134.0
h=3F	0.643	506.0	0.846**	66.0	0.556	149.0
h=4L	0.581	220.0	0.800*	36.0	0.400	72.0
h=4F	0.548	193.0	0.700	35.0	0.500	71.0
h=5L	0.583	89.0	0.571	6.0	0.563	38.0
h=5F	0.375	74.0	0.143	2.0	0.313	31.0
h=6L	0.300	27.0**	0.000	0.0	0.308	17.0
h=6F	0.150**	12.0***	0.000	0.0	0.154	7.0**
HEB6 (Last)						
h=0L	0.450	3150.0**	0.318**	149.0***	0.518	1964.0
h=0F	0.434	3049.0***	0.318**	140.0***	0.494	1956.0
h=1L	0.391**	2553.0***	0.326**	138.0***	0.417	1476.0
h=1F	0.438	2717.0***	0.372	142.0***	0.464	1612.0
h=2L	0.333**	511.0***	0.190**	20.0***	0.561	320.0
h=2F	0.397	579.0***	0.286	27.0***	0.439	342.0
h=3L	0.381	237.0**	0.231	8.0**	0.407	134.0
h=3F	0.381	251.0**	0.231	8.0**	0.556	149.0
h=4L	0.355	97.0**	0.100*	1.0**	0.400	72.0
h=4F	0.355	95.0**	0.100*	1.0**	0.500	71.0
h=5L	0.333	44.0**	0.000*	0.0*	0.563	38.0
h=5F	0.292	36.0***	0.000*	0.0*	0.313	31.0
h=6L	0.250	24.0**	0.000	0.0	0.308	17.0
h=6F	0.150**	12.0***	0.000	0.0	0.154	7.0**

TABLE 15—TESTS OF MEDIAN FORECAST BIAS: SURPLUS

Forecast Horizon	Full Sample		Pre-1991Q1		Post-1990Q4	
	Sign	Signed-Rank	Sign	Signed-Rank	Sign	Signed-Rank
(Last)						
h=0L	0.552	6634.0***	0.369**	892.0	0.700***	2508.0***
h=0F	0.490	5979.0	0.369**	811.0*	0.588*	2329.0***
h=1L	0.458	4994.0	0.375*	752.0*	0.532	1812.0
h=1F	0.444	4845.0	0.406	728.0**	0.481	1755.0
h=2L	0.451	1039.0	0.355	126.0**	0.538	365.0
h=2F	0.423	1029.0	0.323	131.0**	0.513	368.0
h=3L	0.362	393.0	0.250	45.0*	0.400	138.0
h=3F	0.340	404.0	0.200**	45.0*	0.440	131.0
h=4L	0.371	210.0	0.333	26.0	0.368	66.0
h=4F	0.314	194.0	0.267	25.0	0.316	62.0
h=5L	0.357	106.0	0.250	9.0*	0.333	31.0
h=5F	0.286	81.0**	0.167	3.0**	0.333	30.0
h=6L	0.261	40.0**	0.111	1.0*	0.333	22.0
h=6F	0.130***	15.0***	0.000**	0.0**	0.167	7.0*
(First)						
h=0L	0.517	5368.0	0.508	1069.0	0.525	1676.0
h=0F	0.476	4906.0	0.492	960.0	0.463	1542.0
h=1L	0.424*	4155.0**	0.453	853.0	0.405	1235.0*
h=1F	0.424*	4141.0**	0.453	811.0	0.405	1275.0
h=2L	0.408	867.0**	0.419	141.0*	0.410	305.0
h=2F	0.408	864.0**	0.419	140.0*	0.385	299.0
h=3L	0.383	384.0	0.350	51.0	0.400	116.0
h=3F	0.362	400.0	0.350	55.0	0.360	117.0
h=4L	0.343	195.0	0.333	29.0	0.263	60.0
h=4F	0.314	188.0	0.333	27.0	0.263	60.0
h=5L	0.286	100.0*	0.167	8.0*	0.333	25.0
h=5F	0.250*	80.0**	0.167	3.0**	0.267	26.0
h=6L	0.217*	34.0***	0.111	1.0*	0.250	17.0
h=6F	0.174**	16.0***	0.111	1.0*	0.167	7.0*
(1 Yr.)						
h=0L	0.524	5822.0	0.385*	853.0	0.637***	2137.0**
h=0F	0.503	5354.0	0.385*	833.0	0.600*	1957.0
h=1L	0.465	4521.0	0.406	765.0*	0.519	1538.0
h=1F	0.444	4423.0	0.406	741.0**	0.481	1519.0
h=2L	0.437	934.0*	0.387	132.0**	0.410	328.0
h=2F	0.366*	891.0*	0.258**	119.0**	0.385	318.0
h=3L	0.383	386.0	0.300	50.0	0.360	123.0
h=3F	0.362	407.0	0.250	50.0	0.360	117.0
h=4L	0.314	196.0	0.267	27.0	0.316	62.0
h=4F	0.314	191.0	0.333	25.0	0.263	60.0
h=5L	0.321	105.0	0.167	8.0*	0.333	28.0
h=5F	0.250*	82.0**	0.083**	2.0**	0.333	28.0
h=6L	0.261	37.0**	0.111	1.0*	0.250	20.0
h=6F	0.130***	15.0***	0.000**	0.0**	0.167	7.0*
(Pre-B)						
h=0L	0.462	5041.0	0.385*	810.0*	0.525	1818.0
h=0F	0.434	4709.0	0.369**	761.0**	0.487	1705.0
h=1L	0.424*	4142.0**	0.359**	724.0**	0.481	1408.0
h=1F	0.417*	4130.0**	0.406	718.0**	0.430	1401.0
h=2L	0.380	882.0**	0.290*	113.0**	0.359	312.0
h=2F	0.366*	869.0**	0.290*	114.0**	0.385	324.0
h=3L	0.383	355.0*	0.300	41.0**	0.400	122.0
h=3F	0.362	386.0	0.200**	42.0*	0.400	121.0
h=4L	0.314	190.0	0.267	21.0	0.263	60.0
h=4F	0.343	187.0	0.267	23.0	0.263	60.0
h=5L	0.321	100.0*	0.250	9.0*	0.333	29.0
h=5F	0.286	81.0**	0.167	3.0**	0.267	28.0
h=6L	0.217*	36.0**	0.111	1.0*	0.250	19.0
h=6F	0.174**	16.0***	0.111	1.0*	0.167	7.0*

Table 16—Tests of Median Forecast Bias: SURPLUS

Forecast Horizon	Full Sample to 2006		Pre-1986Q1		1986Q1–2006Q4	
	Sign	Signed-Rank	Sign	Signed-Rank	Sign	Signed-Rank
Last						
h=0L	0.512	4874.0	0.386	416.0	0.576	2395.0**
h=0F	0.473	4429.0	0.341**	353.0*	0.541	2288.0**
h=1L	0.438	3673.0	0.349*	343.0	0.476	1729.0
h=1F	0.430	3608.0	0.395	323.0*	0.440	1742.0
h=2L	0.429	720.0*	0.381	57.0*	0.463	353.0
h=2F	0.381	737.0	0.286	60.0	0.415	354.0
h=3L	0.333	270.0*	0.154*	18.0	0.370	131.0
h=3F	0.310*	278.0*	0.154*	20.0	0.370	131.0
h=4L	0.355	142.0	0.300	11.0	0.350	65.0
h=4F	0.290	134.0*	0.200	9.0	0.300	63.0
h=5L	0.333	67.0*	0.143	1.0	0.313	34.0
h=5F	0.250	45.0**	0.000*	0.0*	0.313	31.0
h=6L	0.200*	22.0**	0.000	0.0	0.231	16.0
h=6F	0.100***	9.0***	0.000	0.0	0.154	7.0**
First						
h=0L	0.504	4081.0	0.500	462.0	0.506	1812.0
h=0F	0.481	3785.0	0.432	376.0	0.506	1808.0
h=1L	0.398**	3113.0**	0.395	360.0	0.393*	1303.0**
h=1F	0.414*	3067.0**	0.395	331.0*	0.417	1359.0*
h=2L	0.381	614.0**	0.381	59.0*	0.366	281.0
h=2F	0.381	638.0**	0.381	61.0	0.366	292.0
h=3L	0.357	262.0*	0.231	17.0	0.407	125.0
h=3F	0.333	273.0*	0.308	21.0	0.333	124.0
h=4L	0.323	132.0*	0.300	9.0	0.300	61.0
h=4F	0.290	126.0*	0.200	9.0	0.300	60.0
h=5L	0.250	61.0*	0.143	2.0	0.313	32.0
h=5F	0.208**	44.0**	0.000*	0.0*	0.250	29.0
h=6L	0.150**	19.0***	0.000	0.0	0.154	13.0
h=6F	0.150**	10.0***	0.000	0.0	0.154	7.0**
1 Yr.						
h=0L	0.504	4276.0	0.364*	366.0	0.576	2105.0
h=0F	0.473	3971.0	0.364*	337.0*	0.529	2010.0
h=1L	0.445	3307.0*	0.372	332.0*	0.476	1513.0
h=1F	0.438	3267.0**	0.349*	309.0**	0.476	1560.0
h=2L	0.413	651.0**	0.429	64.0	0.415	299.0
h=2F	0.333**	640.0**	0.238*	57.0*	0.390	303.0
h=3L	0.357	261.0*	0.154*	19.0	0.333	125.0
h=3F	0.310*	277.0*	0.154*	20.0	0.333	130.0
h=4L	0.258**	132.0*	0.200	10.0	0.250	59.0
h=4F	0.290	129.0*	0.200	9.0	0.300	60.0
h=5L	0.292	66.0*	0.143	1.0	0.313	33.0
h=5F	0.208**	46.0**	0.000*	0.0*	0.313	32.0
h=6L	0.200*	20.0***	0.000	0.0	0.231	14.0
h=6F	0.100***	9.0***	0.000	0.0	0.154	7.0**
Pre-B.						
h=0L	0.442	3677.0	0.364*	354.0*	0.482	1770.0
h=0F	0.411*	3534.0	0.341**	314.0**	0.447	1775.0
h=1L	0.391**	3049.0**	0.326**	305.0**	0.417	1390.0*
h=1F	0.398**	3014.0***	0.349*	292.0**	0.417	1411.0*
h=2L	0.333**	608.0**	0.286	47.0**	0.366	305.0
h=2F	0.349**	631.0**	0.286	53.0*	0.390	310.0
h=3L	0.357	236.0**	0.154*	14.0*	0.370	123.0
h=3F	0.310*	258.0**	0.154*	17.0	0.370	127.0
h=4L	0.290	119.0**	0.200	7.0	0.250	59.0
h=4F	0.323	124.0*	0.200	7.0	0.300	61.0
h=5L	0.292	61.0*	0.143	1.0	0.313	33.0
h=5F	0.250	45.0**	0.000*	0.0*	0.250	31.0
h=6L	0.150**	19.0***	0.000	0.0	0.154	13.0
h=6F	0.150**	10.0***	0.000	0.0	0.154	7.0**

TABLE 17—TESTS OF MEDIAN FORECAST BIAS: EXPEND

Forecast Horizon	Full Sample		Pre-1991Q1		Post-1990Q4	
	Sign	Signed-Rank	Sign	Signed-Rank	Sign	Signed-Rank
(Last)						
h=0L	0.538	5975.0	0.554	1347.0*	0.525	1656.0
h=0F	0.566*	6304.0**	0.569	1387.0**	0.563	1772.0
h=1L	0.514	5182.0	0.547	1129.0	0.481	1405.0
h=1F	0.528	5291.0	0.547	1129.0	0.506	1481.0
h=2L	0.408	921.0*	0.548	206.0	0.385	255.0
h=2F	0.408	890.0*	0.581	204.0	0.410	249.0*
h=3L	0.404	376.0	0.550	91.0	0.360	89.0
h=3F	0.426	360.0*	0.550	79.0	0.360	95.0
h=4L	0.200***	63.0***	0.067***	1.0***	0.263	40.0
h=4F	0.200***	62.0***	0.067***	1.0***	0.263	41.0
h=5L	0.214**	48.0***	0.167	3.0**	0.267	21.0
h=5F	0.179***	27.0***	0.083**	1.0**	0.267	20.0
h=6L	0.087***	13.0***	0.000**	0.0**	0.167	9.0
h=6F	0.043***	4.0***	0.000**	0.0**	0.000***	0.0**
(First)						
h=0L	0.607***	6457.0**	0.538	1178.0	0.662***	2149.0**
h=0F	0.641***	6716.0***	0.600*	1274.0	0.675***	2192.0***
h=1L	0.542	5389.0	0.484	1009.0	0.582	1671.0
h=1F	0.521	5480.0	0.469	1015.0	0.557	1737.0
h=2L	0.437	944.0	0.452	200.0	0.564	277.0
h=2F	0.408	907.0*	0.452	194.0	0.385	273.0
h=3L	0.383	335.0**	0.600	79.0	0.360	92.0
h=3F	0.404	360.0*	0.550	79.0	0.360	96.0
h=4L	0.171***	64.0***	0.000***	0.0***	0.316	39.0*
h=4F	0.200***	71.0***	0.067***	1.0***	0.316	39.0*
h=5L	0.214**	49.0***	0.167	4.0**	0.267	22.0
h=5F	0.143***	26.0***	0.000***	0.0**	0.267	21.0
h=6L	0.087***	10.0***	0.000**	0.0**	0.167	8.0*
h=6F	0.043***	5.0***	0.000**	0.0**	0.000***	0.0**
(1 Yr.)						
h=0L	0.621***	6821.0***	0.754***	1663.0***	0.512	1738.0
h=0F	0.614***	6915.0***	0.708***	1589.0***	0.537	1828.0
h=1L	0.556	5722.0	0.641**	1322.0*	0.481	1484.0
h=1F	0.569*	5715.0	0.625**	1271.0	0.519	1532.0
h=2L	0.507	1037.0	0.581	294.0	0.410	266.0
h=2F	0.451	1004.0	0.645	290.0	0.410	262.0
h=3L	0.426	389.0	0.600	111.0	0.360	91.0
h=3F	0.426	372.0	0.550	87.0	0.400	98.0
h=4L	0.229***	75.0***	0.067***	2.0***	0.263	40.0
h=4F	0.229***	68.0***	0.067***	1.0***	0.263	41.0
h=5L	0.214**	54.0***	0.167	5.0**	0.267	21.0
h=5F	0.179***	31.0***	0.083**	1.0**	0.267	22.0
h=6L	0.087***	12.0***	0.000**	0.0**	0.167	10.0
h=6F	0.043***	4.0***	0.000**	0.0**	0.000***	0.0**
(Pre-B)						
h=0L	0.579**	6101.0	0.692***	1598.0***	0.487	1417.0
h=0F	0.600**	6430.0**	0.692***	1559.0***	0.525	1572.0
h=1L	0.514	5265.0	0.563	1279.0	0.468	1268.0
h=1F	0.535	5346.0	0.547	1222.0	0.519	1365.0
h=2L	0.451	982.0	0.581	294.0	0.385	238.0*
h=2F	0.408	951.0	0.581	280.0	0.359	231.0*
h=3L	0.404	371.0	0.550	101.0	0.360	83.0*
h=3F	0.383	359.0*	0.500	89.0	0.320	89.0
h=4L	0.200***	65.0***	0.000***	0.0***	0.263	37.0*
h=4F	0.229***	62.0***	0.067***	1.0***	0.263	40.0
h=5L	0.214**	51.0***	0.167	5.0**	0.267	20.0
h=5F	0.179***	30.0***	0.083**	1.0**	0.267	22.0
h=6L	0.087***	12.0***	0.000**	0.0**	0.167	9.0
h=6F	0.043***	4.0***	0.000**	0.0**	0.000***	0.0**

Table 18—Tests of Median Forecast Bias: EXPEND

Forecast Horizon	Full Sample to 2006		Pre-1986Q1		1986Q1–2006Q4	
	Sign	Signed-Rank	Sign	Signed-Rank	Sign	Signed-Rank
h=0L	0.550	5010.0*	0.545	659.0*	0.553	2060.0
h=0F	0.589**	5261.0**	0.614*	671.0**	0.576	2164.0
h=1L	0.555	4412.0	0.581	566.0	0.548	1804.0
h=1F	0.563	4444.0	0.581	572.0	0.560	1819.0
h=2L	0.556	802.0	0.571	131.0	0.537	328.0
h=2F	0.556	779.0	0.571	101.0	0.537	319.0
h=3L	0.548	332.0	0.615	39.0	0.519	126.0
h=3F	0.476	324.0	0.538	36.0	0.481	137.0
h=4L	0.194***	44.0***	0.000***	0.0**	0.300	38.0**
h=4F	0.194***	43.0***	0.000***	0.0**	0.300	42.0*
h=5L	0.250	36.0***	0.143	1.0	0.250	21.0*
h=5F	0.208**	21.0***	0.143	1.0	0.250	17.0**
h=6L	0.100***	9.0***	0.000	0.0	0.154	8.0*
h=6F	0.050***	3.0***	0.000	0.0	0.077**	2.0**
First						
h=0L	0.620***	5332.0***	0.523	500.0	0.671***	2573.0***
h=0F	0.667***	5562.0***	0.568	562.0	0.718***	2621.0***
h=1L	0.570*	4567.0	0.512	467.0	0.607**	2112.0
h=1F	0.547	4600.0	0.512	486.0	0.571	2111.0
h=2L	0.571	808.0	0.476	98.0	0.610	343.0
h=2F	0.444	778.0	0.524	92.0	0.561	335.0
h=3L	0.548	303.0	0.615	36.0	0.519	126.0
h=3F	0.429	331.0	0.538	35.0	0.407	139.0
h=4L	0.161***	43.0***	0.000***	0.0**	0.250	38.0**
h=4F	0.194***	47.0***	0.000***	0.0**	0.300	47.0
h=5L	0.208**	37.0***	0.143	1.0	0.188	21.0*
h=5F	0.167***	20.0***	0.000*	0.0*	0.250	18.0**
h=6L	0.100***	8.0***	0.000	0.0	0.154	7.0**
h=6F	0.050***	4.0***	0.000	0.0	0.077**	3.0**
1 Yr.						
h=0L	0.651***	5783.0***	0.818***	844.0***	0.565	2227.0*
h=0F	0.651***	5816.0***	0.795***	779.0***	0.576	2302.0**
h=1L	0.602**	4895.0*	0.744***	676.0**	0.536	1928.0
h=1F	0.609**	4868.0*	0.721***	639.0**	0.560	1935.0
h=2L	0.556	1124.0	0.667	139.0	0.463	342.0
h=2F	0.587	875.0	0.619	137.0	0.561	341.0
h=3L	0.571	350.0	0.692	48.0	0.519	128.0
h=3F	0.524	339.0	0.615	39.0	0.481	138.0
h=4L	0.194***	51.0***	0.100*	1.0**	0.250	41.0*
h=4F	0.194***	47.0***	0.000***	0.0**	0.300	45.0*
h=5L	0.250	41.0***	0.143	2.0	0.250	24.0
h=5F	0.208**	25.0***	0.143	1.0	0.250	20.0*
h=6L	0.100***	8.0***	0.000	0.0	0.154	8.0*
h=6F	0.050***	3.0***	0.000	0.0	0.077**	2.0**
Pre-B.						
h=0L	0.612***	5283.0**	0.727***	781.0***	0.553	1994.0
h=0F	0.643***	5525.0***	0.750***	751.0***	0.588*	2159.0
h=1L	0.555	4576.0	0.651**	638.0**	0.512	1757.0
h=1F	0.570*	4604.0	0.628*	608.0	0.548	1820.0
h=2L	0.540	851.0	0.571	137.0	0.463	315.0
h=2F	0.556	826.0	0.571	131.0	0.537	307.0
h=3L	0.452	331.0	0.615	41.0	0.407	119.0
h=3F	0.429	326.0	0.538	36.0	0.407	128.0
h=4L	0.161***	45.0***	0.000***	0.0**	0.250	38.0**
h=4F	0.194***	42.0***	0.000***	0.0**	0.300	41.0*
h=5L	0.250	36.0***	0.143	2.0	0.250	20.0*
h=5F	0.208**	22.0***	0.143	1.0	0.250	18.0**
h=6L	0.100***	9.0***	0.000	0.0	0.154	8.0*
h=6F	0.050***	3.0***	0.000	0.0	0.077**	2.0**

TABLE 19—TESTS OF MEDIAN FORECAST BIAS: RECEIPTS

Forecast Horizon	Full Sample		Pre-1991Q1		Post-1990Q4	
	Sign	Signed-Rank	Sign	Signed-Rank	Sign	Signed-Rank
(Last)						
h=0L	0.621***	7011.0***	0.569	1343.0*	0.662***	2207.0***
h=0F	0.586**	6735.0***	0.569	1252.0	0.600*	2193.0***
h=1L	0.535	5393.0	0.531	1004.0	0.532	1680.0
h=1F	0.493	5212.0	0.531	958.0	0.456	1639.0
h=2L	0.437	1126.0	0.419	182.0	0.436	335.0
h=2F	0.437	1135.0	0.419	192.0	0.436	354.0
h=3L	0.426	463.0	0.600	67.0	0.400	140.0
h=3F	0.468	476.0	0.550	59.0	0.520	141.0
h=4L	0.629	141.0**	0.133**	4.0***	0.632	74.0
h=4F	0.314	129.0***	0.133**	3.0***	0.368	72.0
h=5L	0.357	97.0*	0.167	4.0**	0.533	40.0
h=5F	0.321	77.0**	0.083**	1.0**	0.533	41.0
h=6L	0.304	42.0**	0.000**	0.0**	0.583	24.0
h=6F	0.217*	19.0***	0.000**	0.0**	0.250	8.0*
(First)						
h=0L	0.579**	6322.0**	0.554	1237.0	0.600*	1984.0*
h=0F	0.503	6039.0	0.492	1113.0	0.512	1982.0*
h=1L	0.528	4910.0	0.531	924.0	0.519	1518.0
h=1F	0.493	4816.0	0.500	888.0	0.481	1496.0
h=2L	0.437	1029.0	0.419	160.0	0.410	299.0
h=2F	0.423	1031.0	0.419	173.0	0.410	317.0
h=3L	0.340	409.0	0.300	57.0	0.280	110.0
h=3F	0.383	425.0	0.300	48.0*	0.320	119.0
h=4L	0.314	139.0**	0.133**	4.0***	0.368	68.0
h=4F	0.286*	128.0***	0.133**	3.0***	0.368	72.0
h=5L	0.393	94.0*	0.167	4.0**	0.333	34.0
h=5F	0.321	74.0**	0.083**	1.0**	0.333	32.0
h=6L	0.304	42.0**	0.000**	0.0**	0.500	23.0
h=6F	0.217*	19.0***	0.000**	0.0**	0.250	8.0*
(1 Yr.)						
h=0L	0.648***	7225.0***	0.646**	1489.0***	0.650***	2179.0***
h=0F	0.600**	6929.0***	0.600*	1347.0*	0.600*	2199.0***
h=1L	0.563	5699.0	0.578	1129.0	0.544	1717.0
h=1F	0.521	5557.0	0.547	1072.0	0.494	1711.0
h=2L	0.535	1234.0	0.484	198.0	0.564	353.0
h=2F	0.479	1247.0	0.484	213.0	0.462	376.0
h=3L	0.553	488.0	0.700	75.0	0.560	141.0
h=3F	0.532	519.0	0.600	70.0	0.560	184.0
h=4L	0.657	158.0**	0.133**	5.0***	0.632	80.0
h=4F	0.371	141.0**	0.133**	3.0***	0.579	82.0
h=5L	0.429	107.0	0.167	5.0**	0.600	40.0
h=5F	0.393	87.0**	0.083**	1.0**	0.600	39.0
h=6L	0.348	47.0**	0.000**	0.0**	0.583	24.0
h=6F	0.217*	19.0***	0.000**	0.0**	0.250	9.0
(Pre-B)						
h=0L	0.607***	6295.0**	0.631**	1443.0**	0.588*	1720.0
h=0F	0.531	5920.0	0.554	1254.0	0.512	1753.0
h=1L	0.500	4804.0	0.563	1045.0	0.443	1330.0
h=1F	0.507	4763.0	0.578	1010.0	0.443	1353.0
h=2L	0.408	1053.0	0.452	190.0	0.359	294.0
h=2F	0.423	1044.0	0.516	193.0	0.385	317.0
h=3L	0.362	414.0	0.350	65.0	0.320	116.0
h=3F	0.447	415.0	0.550	56.0	0.360	122.0
h=4L	0.343	141.0**	0.133**	4.0***	0.316	69.0
h=4F	0.314	127.0***	0.133**	3.0***	0.368	70.0
h=5L	0.357	95.0*	0.167	5.0**	0.333	31.0
h=5F	0.321	77.0**	0.083**	1.0**	0.333	31.0
h=6L	0.304	41.0**	0.000**	0.0**	0.333	24.0
h=6F	0.217*	19.0***	0.000**	0.0**	0.250	8.0*

TABLE 20—TESTS OF MEDIAN FORECAST BIAS: RECEIPTS

Forecast Horizon	Full Sample to 2006		Pre-1986Q1		1986Q1–2006Q4	
	Sign	Signed-Rank	Sign	Signed-Rank	Sign	Signed-Rank
Last						
h=0L	0.605**	5339.0***	0.477	589.0	0.671***	2380.0**
h=0F	0.574*	5183.0**	0.523	542.0	0.600*	2407.0**
h=1L	0.531	4151.0	0.558	471.0	0.524	1792.0
h=1F	0.500	3998.0	0.535	436.0	0.488	1766.0
h=2L	0.429	869.0	0.429	77.0	0.415	357.0
h=2F	0.444	879.0	0.429	83.0	0.463	384.0
h=3L	0.405	339.0	0.308	22.0	0.556	158.0
h=3F	0.452	349.0	0.308	19.0	0.556	163.0
h=4L	0.323	90.0***	0.100*	2.0**	0.650	70.0
h=4F	0.290	80.0***	0.100*	1.0**	0.550	66.0
h=5L	0.333	64.0*	0.143	2.0	0.625	43.0
h=5F	0.292	47.0**	0.000*	0.0*	0.563	35.0
h=6L	0.300	25.0**	0.000	0.0	0.385	20.0
h=6F	0.200*	13.0***	0.000	0.0	0.231	9.0*
First						
h=0L	0.566	4962.0*	0.500	523.0	0.600*	2278.0**
h=0F	0.496	4769.0	0.455	450.0	0.518	2312.0**
h=1L	0.516	3851.0	0.535	427.0	0.512	1692.0
h=1F	0.477	3670.0	0.512	394.0	0.464	1629.0
h=2L	0.429	787.0	0.381	66.0	0.415	320.0
h=2F	0.413	794.0	0.381	69.0	0.415	335.0
h=3L	0.310*	302.0	0.231	17.0	0.296	130.0
h=3F	0.357	300.0	0.231	14.0*	0.370	140.0
h=4L	0.290	89.0***	0.100*	2.0**	0.550	69.0
h=4F	0.258**	80.0***	0.100*	1.0**	0.350	66.0
h=5L	0.375	62.0*	0.143	2.0	0.313	36.0
h=5F	0.292	45.0**	0.000*	0.0*	0.313	32.0
h=6L	0.250	24.0**	0.000	0.0	0.385	20.0
h=6F	0.200*	13.0***	0.000	0.0	0.231	9.0*
1 Yr.						
h=0L	0.643***	5672.0***	0.636**	703.0**	0.647***	2404.0**
h=0F	0.597**	5464.0***	0.591	620.0	0.600*	2419.0***
h=1L	0.563	4489.0	0.628*	555.0	0.536	1877.0
h=1F	0.523	4343.0	0.558	511.0	0.512	1868.0
h=2L	0.524	948.0	0.571	89.0	0.537	390.0
h=2F	0.492	983.0	0.524	95.0	0.488	449.0
h=3L	0.571	364.0	0.692	27.0	0.593	161.0
h=3F	0.548	389.0	0.615	24.0	0.593	164.0
h=4L	0.645	103.0**	0.100*	2.0**	0.650	78.0
h=4F	0.355	88.0***	0.100*	1.0**	0.600	74.0
h=5L	0.542	74.0	0.143	2.0	0.625	42.0
h=5F	0.375	54.0**	0.000*	0.0*	0.625	39.0
h=6L	0.300	26.0**	0.000	0.0	0.615	20.0
h=6F	0.200*	13.0***	0.000	0.0	0.231	9.0*
Pre-B.						
h=0L	0.620***	5054.0**	0.568	657.0*	0.647***	2084.0
h=0F	0.543	4776.0	0.523	557.0	0.553	2099.0
h=1L	0.508	3842.0	0.581	497.0	0.464	1522.0
h=1F	0.508	3712.0	0.558	464.0	0.476	1508.0
h=2L	0.397	785.0	0.429	76.0	0.341	299.0
h=2F	0.429	823.0	0.381	76.0	0.415	323.0
h=3L	0.333	303.0	0.308	22.0	0.333	129.0
h=3F	0.429	294.0	0.308	17.0	0.370	133.0
h=4L	0.323	90.0***	0.100*	2.0**	0.400	70.0
h=4F	0.290	80.0***	0.100*	1.0**	0.550	66.0
h=5L	0.333	61.0*	0.143	2.0	0.313	33.0
h=5F	0.292	46.0**	0.000*	0.0*	0.313	32.0
h=6L	0.250	24.0**	0.000	0.0	0.385	19.0
h=6F	0.200*	13.0***	0.000	0.0	0.231	9.0*

TABLE 21—TESTS OF MEDIAN FORECAST BIAS: UNEMP

Forecast Horizon	Full Sample		Pre-1991Q1		Post-1990Q4	
	Sign	Signed-Rank	Sign	Signed-Rank	Sign	Signed-Rank
(Last)						
h=0L	0.772***	6954.5***	0.800***	1540.0***	0.750***	1956.5
h=0F	0.738***	7165.5***	0.754***	1478.5***	0.725***	2155.0**
h=1L	0.694***	6860.0***	0.734***	1433.0***	0.671***	2011.5**
h=1F	0.722***	6939.0***	0.703***	1399.0**	0.747***	2107.5***
h=2L	0.718***	1723.5**	0.774***	362.5**	0.718***	511.0
h=2F	0.704***	1650.5*	0.742***	343.5	0.718***	507.0
h=3L	0.723***	753.0	0.750**	153.5	0.720**	217.5
h=3F	0.723***	731.5	0.800***	165.5*	0.720**	203.5
h=4L	0.771***	446.5	0.867***	101.0*	0.684	120.5
h=4F	0.714**	413.0	0.800**	87.5	0.737*	119.0
h=5L	0.643	185.5	0.583	26.0	0.800**	87.5
h=5F	0.643	155.5	0.333	18.0	0.733	76.5
h=6L	0.391	55.0*	0.111	2.0*	0.750	48.0
h=6F	0.217*	29.0***	0.111	2.0*	0.667	11.5
(First)						
h=0L	0.552	5796.0	0.585	1298.0	0.525	1604.0
h=0F	0.648***	6932.5***	0.662***	1416.5**	0.637***	2098.0**
h=1L	0.611***	6760.0***	0.609*	1368.0**	0.620**	2066.5**
h=1F	0.681***	6741.5***	0.641**	1352.0**	0.722***	2080.5**
h=2L	0.690***	1691.5**	0.742***	358.0*	0.667**	497.5
h=2F	0.704***	1621.5*	0.710**	340.5	0.718***	492.5
h=3L	0.723***	755.5	0.750**	156.0	0.720**	221.0
h=3F	0.702***	722.5	0.750**	164.0*	0.680	208.5
h=4L	0.714**	445.0	0.800**	101.0*	0.737*	122.5
h=4F	0.686*	418.0	0.733	88.0	0.684	120.0
h=5L	0.643	177.0	0.583	25.0	0.800**	89.0
h=5F	0.643	150.5	0.333	18.0	0.733	80.0
h=6L	0.391	53.0*	0.111	2.0*	0.750	48.5
h=6F	0.217*	28.0***	0.111	2.0*	0.667	11.5
(1 Yr.)						
h=0L	0.552	5798.5	0.585	1298.0	0.525	1603.5
h=0F	0.648***	6933.0***	0.662***	1416.5**	0.637***	2097.0**
h=1L	0.611***	6756.5***	0.609*	1368.0**	0.620**	2065.0**
h=1F	0.681***	6737.0***	0.641**	1352.0**	0.722***	2078.5**
h=2L	0.690***	1691.5**	0.742***	358.0*	0.667**	499.5
h=2F	0.704***	1622.5*	0.710**	340.5	0.718***	492.5
h=3L	0.723***	755.5	0.750**	156.0	0.720**	221.0
h=3F	0.702***	722.5	0.750**	164.0*	0.680	208.5
h=4L	0.714**	445.0	0.800**	101.0*	0.737*	122.5
h=4F	0.686*	418.0	0.733	88.0	0.684	120.0
h=5L	0.643	177.0	0.583	25.0	0.800**	89.0
h=5F	0.643	150.5	0.333	18.0	0.733	79.0
h=6L	0.391	53.0*	0.111	2.0*	0.750	48.5
h=6F	0.217*	28.0***	0.111	2.0*	0.667	11.5
(CV)						
h=0L	0.552	5808.0	0.585	1298.0	0.525	1610.0
h=0F	0.648***	6938.0***	0.662***	1416.5**	0.637***	2102.0**
h=1L	0.611***	6766.0***	0.609*	1368.0**	0.620**	2068.0**
h=1F	0.681***	6744.0***	0.641**	1352.0**	0.722***	2082.0**
h=2L	0.690***	1691.5**	0.742***	358.0*	0.692**	497.5
h=2F	0.704***	1621.5*	0.710**	340.5	0.718***	492.5
h=3L	0.723***	755.5	0.750**	156.0	0.720**	221.0
h=3F	0.702***	722.5	0.750**	164.0*	0.680	208.5
h=4L	0.714**	445.0	0.800**	101.0*	0.737*	122.5
h=4F	0.686*	418.0	0.733	88.0	0.737*	121.0
h=5L	0.643	177.0	0.583	25.0	0.800**	89.0
h=5F	0.643	150.5	0.333	18.0	0.733	80.0
h=6L	0.391	53.0*	0.111	2.0*	0.750	48.5
h=6F	0.217*	28.0***	0.111	2.0*	0.667	11.5

Table 22—Tests of Median Forecast Bias: UNEMP

Forecast Horizon	Full Sample to 2006		Pre-1986Q1		1986Q1–2006Q4	
	Sign	Signed-Rank	Sign	Signed-Rank	Sign	Signed-Rank
Last						
h=0L	0.791***	5747.5***	0.773***	680.5**	0.800***	2518.5***
h=0F	0.767***	6045.0***	0.727***	650.0*	0.788***	2799.5***
h=1L	0.734***	5833.5***	0.674**	596.0	0.762***	2732.0***
h=1F	0.758***	5907.0***	0.651**	596.5	0.810***	2802.5***
h=2L	0.746***	1448.5***	0.714*	155.0	0.780***	665.5***
h=2F	0.730***	1400.0**	0.667	148.0	0.780***	654.0***
h=3L	0.762***	667.5**	0.692	61.5	0.815***	299.5**
h=3F	0.762***	647.5**	0.769*	69.0	0.778***	288.0*
h=4L	0.806***	386.5**	0.800*	44.0	0.800**	169.5*
h=4F	0.742**	356.0	0.700	35.0	0.800**	167.0*
h=5L	0.625	160.5	0.571	7.0	0.750	96.5
h=5F	0.667	115.5	0.143	1.0	0.750	91.0
h=6L	0.350	38.0*	0.000	0.0	0.692	53.0
h=6F	0.200*	22.0**	0.000	0.0	0.692	16.0
First						
h=0L	0.558	4703.5	0.545	588.0	0.565	1999.5
h=0F	0.667***	5737.0***	0.636**	624.0	0.682***	2622.0***
h=1L	0.641***	5657.0***	0.581	571.5	0.667***	2665.0***
h=1F	0.703***	5698.0***	0.581	578.5	0.762***	2700.0***
h=2L	0.714***	1417.5**	0.667	150.0	0.732***	637.0**
h=2F	0.730***	1381.0**	0.619	148.0	0.780***	636.0**
h=3L	0.762***	665.0**	0.692	62.0	0.815***	294.0**
h=3F	0.738***	639.5*	0.692	67.0	0.778***	285.5*
h=4L	0.742**	385.0**	0.800*	44.0	0.750**	164.0
h=4F	0.710**	361.5	0.700	36.0	0.750**	163.0
h=5L	0.625	141.0	0.286	5.0	0.750	94.5
h=5F	0.667	112.0	0.000*	0.0*	0.750	89.5
h=6L	0.350	36.0*	0.000	0.0	0.692	54.0
h=6F	0.200*	21.0**	0.000	0.0	0.615	16.0
1 Yr.						
h=0L	0.558	4703.5	0.545	588.0	0.565	1999.5
h=0F	0.667***	5737.0***	0.636**	624.0	0.682***	2622.0***
h=1L	0.641***	5657.0***	0.581	571.5	0.667***	2665.0***
h=1F	0.703***	5698.0***	0.581	578.5	0.762***	2700.0***
h=2L	0.714***	1417.5**	0.667	150.0	0.732***	637.0**
h=2F	0.730***	1381.0**	0.619	148.0	0.780***	636.0**
h=3L	0.762***	665.0**	0.692	62.0	0.815***	294.0**
h=3F	0.738***	639.5*	0.692	67.0	0.778***	285.5*
h=4L	0.742**	385.0**	0.800*	44.0	0.750**	164.0
h=4F	0.710**	361.5	0.700	36.0	0.750**	163.0
h=5L	0.625	141.0	0.286	5.0	0.750	94.5
h=5F	0.667	112.0	0.000*	0.0*	0.750	89.5
h=6L	0.350	36.0*	0.000	0.0	0.692	54.0
h=6F	0.200*	21.0**	0.000	0.0	0.615	16.0
CV						
h=0L	0.558	4703.5	0.545	588.0	0.565	1999.5
h=0F	0.667***	5737.0***	0.636**	624.0	0.682***	2622.0***
h=1L	0.641***	5657.0***	0.581	571.5	0.667***	2665.0***
h=1F	0.703***	5698.0***	0.581	578.5	0.762***	2700.0***
h=2L	0.714***	1417.5**	0.667	150.0	0.732***	637.0**
h=2F	0.730***	1381.0**	0.619	148.0	0.780***	636.0**
h=3L	0.762***	665.0**	0.692	62.0	0.815***	294.0**
h=3F	0.738***	639.5*	0.692	67.0	0.778***	285.5*
h=4L	0.742**	385.0**	0.800*	44.0	0.750**	164.0
h=4F	0.710**	361.5	0.700	36.0	0.750**	163.0
h=5L	0.625	141.0	0.286	5.0	0.750	94.5
h=5F	0.667	112.0	0.000*	0.0*	0.750	89.5
h=6L	0.350	36.0*	0.000	0.0	0.692	54.0
h=6F	0.200*	21.0**	0.000	0.0	0.615	16.0