WORKING PAPERS
RESEARCH DEPARTMENT

WORKING PAPER NO. 16-17

DO GDP FORECASTS RESPOND EFFICIENTLY TO CHANGES IN INTEREST RATES?

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

Katherine Marsten
Board of Governors of the Federal Reserve System

June 22, 2016

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

Do GDP Forecasts Respond Efficiently to Changes in Interest Rates?

By Dean Croushore and Katherine Marsten*

June 22, 2016

In this paper, we examine and extend the results of Ball and Croushore (2003) and Rudebusch and Williams (2009), who show that the output forecasts in the Survey of Professional Forecasters (SPF) are inefficient. Ball and Croushore show that the SPF output forecasts are inefficient with respect to changes in monetary policy, as measured by changes in real interest rates, while Rudebusch and Williams show that the forecasts are inefficient with respect to the yield spread. In this paper, we investigate the robustness of both claims of inefficiency, using real-time data and exploring the impact of alternative sample periods on the results.
Keywords: real-time data, output forecasts, yield spread, monetary policy

* Croushore: Robins School of Business, University of Richmond, dcrousho@richmond.edu. Marsten: Federal Reserve Board of Governors, katherine.marsten@frb.gov. We thank the University of Richmond for summer research support. We also thank participants at seminars and conferences at CIREQ, the Society for Economic Measurement, the Western Economic Association, the Federal Reserve Bank of San Francisco, and the University of Richmond for their comments and suggestions.

Forecasters have incentives to create the best forecasts they can. Studies that examine the efficiency of forecasts from surveys, including the Survey of Professional Forecasters (SPF) or the Livingston Survey, such as those by Pesaran and Weale (2006) and Croushore (2010), suggest that such aggregate forecasts are reasonably accurate but not completely efficient. Papers by Ball and Croushore (2003) and Rudebusch and Williams (2009) show that the output forecasts in the SPF are inefficient. Ball and Croushore show that the SPF output forecasts are inefficient with respect to changes in monetary policy (as measured by changes in real interest rates), while Rudebusch and Williams show that the forecasts are inefficient with respect to the yield spread. In this paper, we investigate the robustness of both claims of inefficiency, using real-time data and exploring the impact of alternative sample periods on the results.

I. Research on Forecast Inefficiency

Ball and Croushore show that changes in monetary policy lead professional forecasters to modify their output forecasts, but they do not do so efficiently. The forecasters change their output forecasts in response to a change in monetary policy but not by a sufficient amount. (Ball and Croushore find that the forecasters do respond efficiently in their forecasts for inflation, however.) To illustrate this result, Ball and Croushore compare the SPF output growth forecasts both with actual output growth and with a forecast from a univariate forecasting model, which assumes that output growth follows an AR(1) process with a mean shift in 1973Q2. They then examine the forecast errors for both the time-series model and the SPF, as well as the differences between the time-series model and the SPF forecasts.

To measure monetary policy, Ball and Croushore use the change over the previous year in the real federal funds rate ($FF1$), which is defined as the nominal federal funds rate minus the expected inflation rate over the next year. Because of lags in the effect of monetary policy, they also look at an additional one-year lag of that measure ($FF2$). The forecasters in the SPF would know the values of $FF1$ and $FF2$ at the time they make their forecasts of output growth for the coming year.

Ball and Croushore find that $FF1$ is correlated with the time-series forecast errors, which suggests that monetary policy has an impact on output because it moves output in a manner that differs from what the time-series model suggests. They also find that $FF1$ is correlated with the differences between the time-series forecast and the SPF, which suggests that the SPF participants use information that is different from just a univariate time-series model. However, the SPF forecast errors turn out to be negatively correlated with $FF1$, suggesting that an increase in the real federal funds rate over the past year leads output growth to fall by more than the SPF participants believe it will. Thus, the $FF1$ measure could be used to improve upon the forecasts of the SPF. Ball and Croushore examine the robustness of their results to a more general lag structure, to potential regime shifts, to including output shocks in the regression, and to using changes in the nominal federal funds rate instead of the real federal funds rate as a measure of monetary policy. Their results are robust to all of these variations.

Rudebusch and Williams focus mainly on the use of the yield spread to predict recessions, comparing a yield spread probit model with the SPF probability forecasts of a recession, and their results are robust, as Croushore and Marsten (2015) show. In the last section of their paper, however, Rudebusch and Williams also examine the forecast errors of SPF real output forecasts and their correlation with the lagged yield spread. They find that the SPF real output growth forecast errors are negatively correlated with the yield spread at many horizons (all except current quarter forecasts), suggesting that the SPF forecasters do not efficiently use information from the yield curve. To some extent, this may be due to the early years of the SPF because their evidence is weaker for a sample that is restricted to data after 1987.

II. Data

The forecast data that are the center of both Ball and Croushore (2003) and Rudebusch and Williams (2009) come from the SPF, which began in 1968 as a joint effort by the American Statistical Association and the National Bureau of Economic Research and was called the Economic Outlook Survey.[1] The Federal Reserve Bank of Philadelphia took over the survey in 1990 and renamed it the Survey of Professional Forecasters.[2] Participants must be professional forecasters actively engaged in the forecasting business who are capable of making quarterly forecasts of numerous macroeconomic variables.

Ball and Croushore examine the SPF forecast of the average growth rate of real output over the coming year. For example, in the 1991Q4 survey, the forecasters provided real output forecasts for each quarter from 1991Q4 to 1992Q4. Ball and Croushore examine the forecasted growth rate over that four-quarter period. More generally, they examine the one-year-ahead forecasts, y_t^e, defined as

$$y_t^e = (\frac{Y_{t+4}^e}{Y_t^e} - 1) \times 100\%, \tag{1}$$

where Y_t^e is the level of the output forecast at date t. They compare the forecast with the actual growth rate over the same period, which is

$$y_t = (\frac{Y_{t+4}}{Y_t} - 1) \times 100\%, \tag{2}$$

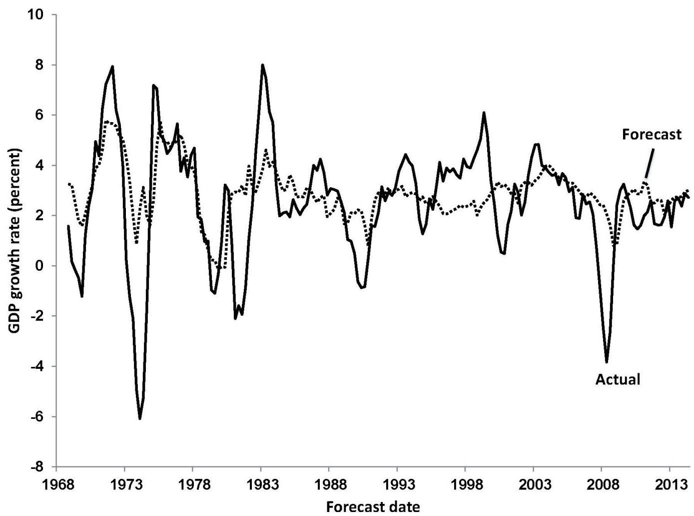
where Y_t is the level of actual output at date t (the definition of "actual" is discussed below).

Figure 1 plots both y_t^e and y_t from 1968Q4 to 2014Q2. Note that the growth

[1]See Zarnowitz and Braun (1993).
[2]See Croushore (1993).

rate forecast is smoother than the actual growth rate, which is a property of optimal forecasts.

Figure 1. Data on GDP Forecast and Actual, Ball–Croushore



Note: The figure shows the forecast for the average growth rate of GDP over the next four quarters with the forecast date shown on the horizontal axis, along with the actual value of GDP growth over those same four quarters. The sample period covers SPF forecasts made from 1968Q4 to 2014Q2.

The forecast error for Ball and Croushore is

(3) $$e_t = y_t - y_t^e.$$

Ball and Croushore run the regression given by this equation[3]

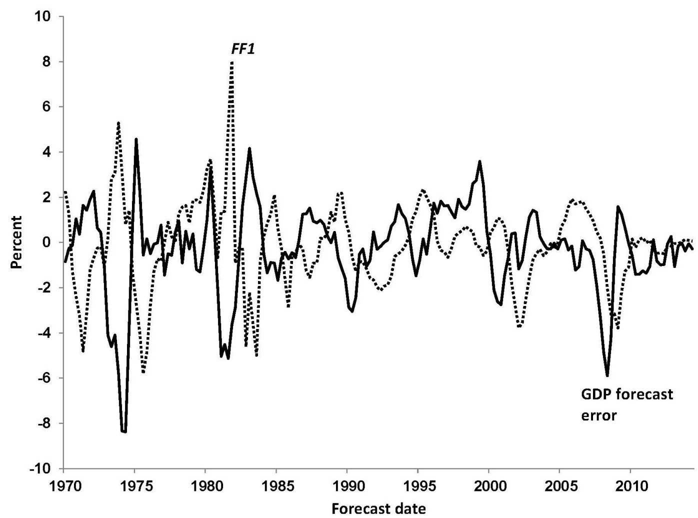
(4) $$e_t = \beta_1 FF1_t + \beta_2 FF2_t + \epsilon_t,$$

where each regression is run twice: once with β_2 set equal to zero and once with both coefficients allowed to be nonzero.

[3]Note that the regressions do not contain constant terms. We also tested all of the regressions used in this paper to see if a constant term was ever statistically significant, and it was not, which suggests that the forecasts are not biased.

Figure 2 plots the forecast error and the $FF1$ variable used in this regression. A negative correlation is readily apparent in the data. Because the $FF1$ variable requires a four-quarter-lagged value of the real federal funds rate (based on the survey inflation expectation from the previous quarter), there is a five-quarter lag for the start of the $FF1$ variable, so the sample period for $FF1$ begins in 1970Q1.

Figure 2. GDP Forecast Error and Measure of Monetary Policy, Ball–Croushore



Note: The figure shows the GDP forecast error over the next four quarters with the forecast date shown on the horizontal axis, along with the $FF1$ measure of monetary policy known to the forecasters at that date. The sample period covers SPF forecasts made from 1970Q1 to 2014Q2.

Rudebusch and Williams test the SPF forecast for output growth at various horizons. For a forecast made at date t, the forecasters have information on output at date $t-1$ and make forecasts for horizons (h) of 0, 1, 2, 3, and 4 quarters, defined as

$$(5) \qquad y^e_{t+h|t-1} = (((\frac{Y^e_{t+h|t-1}}{Y^e_{t+h-1|t-1}})^4) - 1) \times 100\%,$$

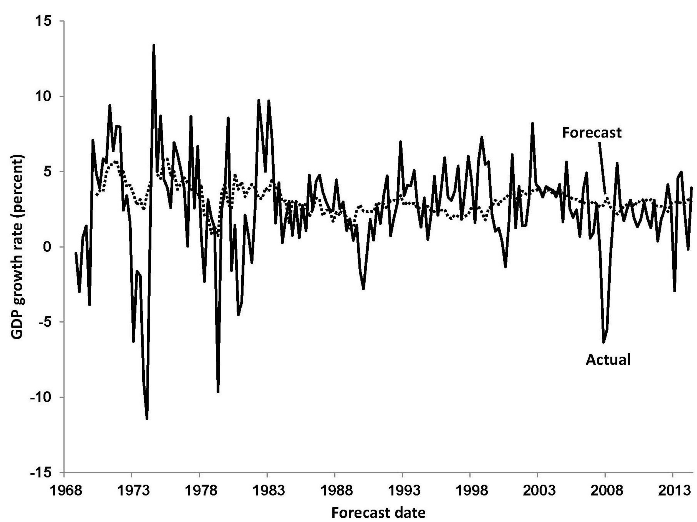
where $h = 0$, 1, 2, 3, and 4, and $Y^e_{t+h|t-1}$ is the level of the output forecast made at date t for date $t+h$, using data on output through date $t-1$. Rudebusch and

Williams test those forecasts against actual values, which are calculated as

$$(6) \qquad y_{t+h} = (((\frac{Y_{t+h}}{Y_{t+h-1}})^4) - 1) \times 100\%.$$

Figure 3 plots both $y^e_{t+4|t-1}$ and y_{t+4} from 1968Q4 to 2014Q2. Note that the four-quarter-ahead forecast is very smooth, especially since the start of the Great Moderation in the early 1980s.

FIGURE 3. DATA ON FOUR-QUARTER-AHEAD GDP FORECAST AND ACTUAL, RUDEBUSCH-WILLIAMS



Note: The figure shows the four-quarter-ahead forecast for the growth rate of GDP with the forecast date shown on the horizontal axis, along with the actual value of GDP growth in that quarter. The sample period covers SPF forecasts made from 1968Q4 to 2014Q2, but note that there are a few missing observations because, in the early years, the SPF did not consistently ask for four-quarter-ahead forecasts.

The h-horizon forecast error is defined as

$$(7) \qquad e_{t+h|t-1} = y_{t+h} - y^e_{t+h|t-1}.$$

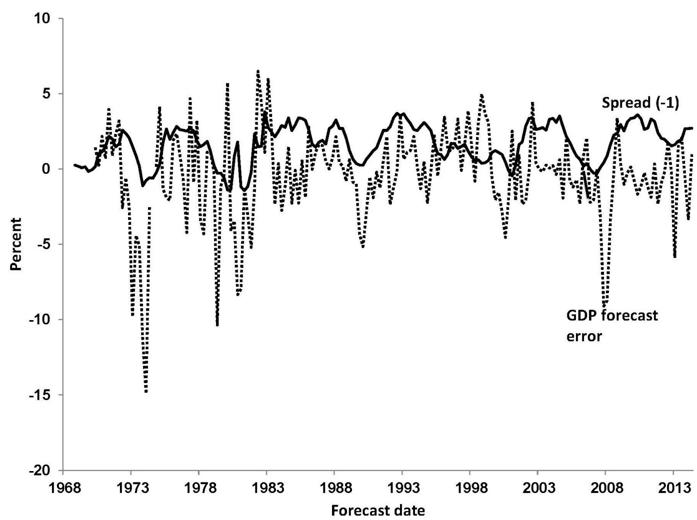
Rudebusch and Williams run the regression

$$e_{t+h|t-1} = \alpha + \beta y^e_{t+h|t-1} + \gamma S_{t-1} + \epsilon_{t+h|t-1}, \tag{8}$$

where S_{t-1} is the average yield spread in the quarter before the survey forecast made at time t. They run the regression three times for every time horizon: once with β and γ both equal to zero, once with γ set equal to zero, and once with all three coefficients allowed to be nonzero.

Figure 4 plots the four-quarter-ahead forecast error and the spread variable used in this regression. The relationship between the two measures is not clear because of the volatility of the forecast error.

Figure 4. Four-Quarter-Ahead GDP Forecast Error and Spread, Rudebusch-Williams



Note: The figure shows the four-quarter-ahead GDP forecast error with the forecast date shown on the horizontal axis, along with the spread variable known to the forecasters at that date. The sample period covers SPF forecasts made from 1968Q4 to 2014Q2, but note that there are a few missing observations because, in the early years, the SPF did not consistently ask for four-quarter-ahead forecasts.

To evaluate the forecasts, a researcher must know the data that were available to the SPF forecasters at the time they made their forecasts. For this purpose, we use the Real-Time Data Set for Macroeconomists (RTDSM), which was created by Croushore and Stark (2001) and made available on the website of the Federal Reserve Bank of Philadelphia. The RTDSM provides information on real output (GNP before 1992, GDP since 1992) and other major macroeconomic variables, as someone standing at the middle of any month from November 1965 to today would have viewed the data. We call the information available to someone at a date in time a "vintage," and the data are known as "real-time data."

Ball and Croushore used a primitive type of real-time data to measure the "actual" value of output in their paper. The RTDSM did not yet exist, so they created a set of "first-final data," which are the data on output released by the government at the end of the third month of the subsequent quarter. For example, for output in the first quarter of the year, the government produces the initial release of the data at the end of April, revises it at the end of May, and releases the first-final data at the end of June. So, all of Ball and Croushore's results were based on using this first-final data release for each period. Similarly, Rudebusch and Williams's research is based on the first-final data, but they also tested the robustness of their results to using the initial release of the data each quarter as well as using the last vintage of data available to them in 2007.

We will use the real-time concepts that the two previous studies used, as well as some additional real-time concepts, in our analysis. As it turns out, much additional information about real output becomes available in the government's annual revision of the data, which is usually released at the end of July each year. So, we will examine the robustness of the results to the use of the annual release of the data, which is a more accurate measure of real output than the initial or first-final releases. For example, Holdren (2014) shows that the percentage of GDP measures coming from "comprehensive data," which are source data that entirely or nearly cover the population, rises from 38.5 percent in the first-final data release to 73.5 percent in the first annual revision (and ultimately to 96.7 percent by the time of a benchmark revision).

Rudebusch and Williams determine the yield spread as the difference between the interest rates on 10-year U.S. government Treasury bonds and three-month Treasury bills. We use the data on these two series from the FRED database maintained by the Federal Reserve Bank of St. Louis, using quarterly averages. Interest rates are not revised, so there is no need for the use of real-time data for the yield spread.

The yield spread can change for a variety of reasons, but one of the most important reasons is because of changes in monetary policy. A change in monetary policy causes current and expected future short-term interest rates to change, thus affecting the yield spread. The yield spread that Rudebusch and Williams use is strongly correlated with the $FF1$ variable used by Ball and Croushore. The simple contemporaneous correlation coefficient between the two measures from 1970 to 2014 is -0.54.[4]

III. Replication Results

We begin our study by replicating the results of Ball and Croushore (2003). Their main result comes from regressing the one-year-ahead forecast error $y_t - y_t^e$ on variable $FF1$ in one regression and on variables $FF1$ and $FF2$ in a second regression. We use (for now) the first-final data for y_t, as in Ball and Croushore. We use the mean forecast across SPF participants for our measure of y_t^e, which is slightly different than Ball and Croushore, who used the median forecast. Theory suggests that the mean forecast is the most appropriate concept to use for aggregating forecasts, as discussed in the literature on forecast encompassing and statistical theory.[5] The results of the replication are shown in Table 1.

As Table 1 shows, we replicate the Ball and Croushore results almost exactly, which suggests that the differences between the mean and median SPF forecasts

[4]In fact, when we use the yield spread used by Rudebusch and Williams in place of the $FF1$ variable in the Ball and Croushore study, we obtain similar results to those from using the $FF1$ variable. Similarly, when we use the $FF1$ variable used by Ball and Croushore in place of the yield spread in the Rudebusch and Williams study, we obtain similar results to those using the yield spread. In both cases, there are some quantitative differences, but they are relatively small.

[5]See Gneiting (2011).

Table 1—Ball–Croushore Results and Replication

Regression: $e_t = \beta_1 FF1_t + \beta_2 FF2_t + \epsilon_t$

Original		
$FF1$	**−0.464**	**−0.466**
	(0.143)	(0.155)
$FF2$		−0.138
		(0.085)
χ^2 sig.	<0.01	<0.01
$\overline{R}^2$	0.20	0.21
Replication		
$FF1$	**−0.489**	**−0.492**
	(0.146)	(0.158)
$FF2$		−0.137
		(0.089)
χ^2 sig.	<0.01	<0.01
$\overline{R}^2$	0.21	0.22

Note: The table shows the original results reported by Ball and Croushore (2003) and our replication using mean SPF forecast data instead of median forecast data. Numbers in parentheses are HAC standard errors to adjust for overlapping observations. Bold numbers indicate coefficients that are statistically significantly different from zero at the 5 percent level. The sample period covers SPF forecasts made from 1968Q4 to 1995Q2, so given the lags in the calculation of $FF1$ and $FF2$, the regression is run on data from forecasts made from 1970Q1 to 1995Q2 for the first regression and 1971Q1 to 1995Q2 for the second regression containing $FF2$.

are very small indeed. This fact can also be seen by plotting the data, which we do not show here to conserve space. Our replication confirms the Ball and Croushore result that the real output growth forecast error, using first-final data as actuals, is negatively correlated with the lagged change in the real federal funds rate, $FF1$. So, the forecasters should have been able to make better output growth forecasts by using the data on monetary policy changes more effectively.

We perform the same type of replication exercise for the study by Rudebusch and Williams. We regress the h-horizon forecast error for output growth (where $h = 0, 1, 2, 3,$ and 4) on a constant, the SPF forecast of output growth, and the yield spread from the quarter before the quarter in which the SPF forecast is made, so that it reflects the information available to the forecasters at the time they made their forecasts. In Table 2, we compare our results with those reported by Rudebusch and Williams. Bold numbers indicate coefficients that are significantly different from zero at the 5 percent level. HAC standard errors are used to account for overlapping observations but are not shown to conserve space. The full sample begins with forecasts made in 1968Q4 and the post-1987 sample begins with forecasts made in 1988Q1; both samples end with forecasts made in 2007Q1.[6]

Although we are not able to replicate the Rudebusch and Williams results exactly, we broadly confirm their finding that there is significant in-sample fit for the yield spread in explaining SPF output forecast errors. However, the coefficient on the yield spread has the opposite sign of that found by Rudebusch and Williams. We find a positive coefficient, which means that a larger spread is correlated with a larger forecast error. So, the forecasters did not increase their GDP forecasts enough as the spread increased (that is, as long-term interest rates rose relative to short-term interest rates). For every forecast horizon except for the current-quarter forecast, we find a significant coefficient for the yield spread. However, most of the F-tests for the overall regression have p-values that do not reject the null hypothesis that all the coefficients are zero. To confirm this finding, we also run the same type of exercise, leaving out the SPF forecast term, to see if the yield spread alone is significant. This is also important because we would expect the SPF forecast to be correlated with the yield spread, so these two terms on the right-hand side of the original regression equation are not likely to be independent. The results of dropping the SPF forecast from the regression are shown in Table 3.

The results show that there is some evidence that the yield spread itself is significant in explaining real GDP forecast errors, at least in the full sample. In Table 3, we see that the coefficient on the yield spread is significantly different from zero for the three- and four-quarter horizons for the full sample period. However, the evidence suggests that the relationship has changed and that since 1987, the forecasters no longer respond inefficiently to changes in the yield spread.

[6] Because of missing observations, the full-sample four-quarter-ahead forecast sample begins in 1970Q2.

Table 2—Rudebusch–Williams Results and Replication

Regression: $e_{t+h|t-1} = \alpha + \beta y^e_{t+h|t-1} + \gamma S_{t-1} + \epsilon_{t+h|t-1}$

	Full Sample			Post-1987 Sample		
Original						
			Current-quarter forecast			
Constant	0.31	0.06	−0.04	0.47	0.67	0.43
SPF forecast		0.10	0.08		−0.08	−0.10
Yield spread			−0.10			0.16
F-test (p-value)	0.11	0.18	0.34	0.01	0.56	0.51
			One-quarter-ahead forecast			
Constant	−0.01	−0.16	−0.52	0.25	0.41	0.21
SPF forecast		0.05	−0.19		−0.06	−0.14
Yield spread			**−0.65**			−0.23
F-test (p-value)	0.96	0.66	0.01	0.32	0.80	0.48
			Two-quarter-ahead forecast			
Constant	−0.24	0.21	−0.14	0.19	0.82	0.84
SPF forecast		−0.15	**−0.50**		−0.24	−0.43
Yield spread			**−0.88**			−0.28
F-test (p-value)	0.47	0.40	0.00	0.53	0.43	0.34
			Three-quarter-ahead forecast			
Constant	−0.50	−0.13	−0.70	0.08	1.40	1.47
SPF forecast		−0.11	−0.31		−0.47	**−0.71**
Yield spread			**−0.76**			−0.32
F-test (p-value)	0.18	0.63	0.02	0.81	0.19	0.16
			Four-quarter-ahead forecast			
Constant	−0.41	0.54	−0.33	0.06	1.55	1.24
SPF forecast		−0.29	−0.37		−0.53	**−0.69**
Yield spread			**−0.68**			−0.43
F-test (p-value)	0.29	0.19	0.00	0.87	0.13	0.05
Replication						
			Current-quarter forecast			
Constant	0.36	0.03	−0.08	**0.49**	0.72	0.46
SPF forecast		0.14	0.11		−0.09	−0.11
Yield spread			0.11			0.17
F-test (p-value)	0.05	0.02	0.06	0.01	0.06	0.01
			One-quarter-ahead forecast			
Constant	0.08	−0.02	−0.32	0.29	0.50	0.37
SPF forecast		0.04	−0.21		−0.08	−0.15
Yield spread			**0.63**			0.18
F-test (p-value)	0.77	0.88	0.18	0.29	0.58	0.13
			Two-quarter-ahead forecast			
Constant	−0.13	−0.20	−0.38	0.22	0.29	0.32
SPF forecast		0.02	−0.27		−0.03	−0.15
Yield spread			0.66			0.17
F-test (p-value)	0.71	0.93	0.25	0.53	0.75	0.10
			Three-quarter-ahead forecast			
Constant	−0.34	0.51	0.24	0.10	1.73	1.89
SPF forecast		−0.27	−0.55		−0.58	−0.81
Yield spread			**0.72**			0.29
F-test (p-value)	0.39	0.55	0.06	0.80	0.56	0.13
			Four-quarter-ahead forecast			
Constant	−0.36	−0.11	−1.05	0.09	1.89	1.89
SPF forecast		−0.08	−0.16		−0.64	**−0.89**
Yield spread			**0.74**			0.40
F-test (p-value)	0.42	0.70	0.03	0.81	0.45	<0.01

Note: The table shows the original results reported by Rudebusch and Williams (2009) and our replication. Bold numbers indicate coefficients that are statistically significantly different from zero at the 5 percent level. HAC standard errors are used to account for overlapping observations but are not shown to conserve space. The full sample begins with forecasts made in 1968Q4 (except for the four-quarter-ahead horizon case, when the sample begins in 1970Q2), and the post-1987 sample begins with forecasts made in 1988Q1; both samples end with forecasts made in 2007Q1.

Table 3—Rudebusch–Williams Results with Yield Spread Alone

Regression: $e_{t+h|t-1} = \alpha + \gamma S_{t-1} + \epsilon_{t+h|t-1}$

	Full Sample	Post-1987 Sample
	Current-quarter forecast	
Constant	0.05	0.22
Yield spread	0.20	0.15
F-test (p-value)	0.04	<0.01
	One-quarter-ahead forecast	
Constant	−0.63	0.03
Yield spread	0.46	0.15
F-test (p-value)	0.16	0.07
	Two-quarter-ahead forecast	
Constant	−0.93	−0.02
Yield spread	0.51	0.14
F-test (p-value)	0.19	0.23
	Three-quarter-ahead forecast	
Constant	−1.24	−0.11
Yield spread	**0.58**	0.12
F-test (p-value)	0.05	0.56
	Four-quarter-ahead forecast	
Constant	−1.55	−0.36
Yield spread	**0.73**	0.26
F-test (p-value)	0.02	0.15

Note: The table shows the rationality test used by Rudebusch and Williams (2009) but with only the yield spread used in the regression. Bold numbers indicate coefficients that are statistically significantly different from zero at the 5 percent level. HAC standard errors are used to account for overlapping observations but are not shown to conserve space. The full sample begins with forecasts made in 1968Q4, and the post-1987 sample begins with forecasts made in 1988Q1; both samples end with forecasts made in 2007Q1.

IV. Robustness Exercises

A. Extending the Sample

We now have 18 years of additional data since Ball and Croushore ran their research and eight years of additional data since Rudebusch and Williams ran theirs. Furthermore, the past seven years include some very difficult times for forecasters, with a deep recession followed by a very weak recovery. So, we update the data to see if the same results hold true for both studies.

However, in examining the SPF data carefully, we discovered a potentially serious problem in the survey design in the early years of the survey. The GNP deflator forecasts in the early years were rounded to the nearest whole number, which introduces substantial volatility in the forecasts. For example, forecaster 21 in the 1968Q4 survey forecast the following pattern for the deflator: 122 in 1968Q3, 123 in 1968Q4, 125 in 1969Q1, 126 in 1969Q2, 127 in 1969Q3, and 128 in 1969Q4. This implies a peculiar pattern in the inflation forecasts: 3.3 percent in 1968Q4, 6.7 percent in 1969Q1, and 3.2 percent in 1969Q2, 1969Q3, and 1969Q4. The fact that the forecasters are rounding makes it appear that there is a temporary jump in inflation in 1969Q1, which is simply a feature of the lack of precision in the survey answers. Because the SPF did not explicitly ask for real output forecasts in the early years, the real output forecasts are derived from nominal output forecasts and the deflator. This lack of precision is especially damaging for the Rudebusch and Williams study, which uses quarterly growth rates and thus is affected more strongly by the lack of precision in the deflator forecasts. For the Ball and Croushore study, which uses forecasts over a full year, the lack of precision should matter a bit less. Nonetheless, to keep from falsely finding inefficient forecasts solely because of the imprecision of the survey, we will not use any SPF surveys made before 1970Q4, which is when the SPF finally added a decimal place to the deflator forecasts.

Table 4 shows the results of extending the Ball and Croushore results from Table 1 to 2014Q2 and starting the analysis using SPF surveys beginning in 1970Q4. In addition, the table shows an additional permutation in which we use the nominal federal funds rate instead of the real federal funds rate in measuring monetary policy. This provides an additional test of the robustness of the results.

The results show that the original Ball and Croushore results hold up reasonably well to changing the ending date of the sample from 1995Q2 to 2014Q2, which is an additional 19 years of data. The coefficients on monetary policy become slightly smaller, but the $FF1$ term remains significant in all regressions. Using the nominal interest rate instead of the real interest rate in measuring changes in monetary policy does not matter very much, suggesting that the results are

Table 4—Ball–Croushore Results and Replication, Extended Sample

Regression: $e_t = \beta_1 FF1_t + \beta_2 FF2_t + \epsilon_t$

Interest rate	Real		Nominal	
$FF1$	**−0.375**	**−0.367**	**−0.333**	**−0.321**
	(0.156)	(0.163)	(0.141)	(0.146)
$FF2$		−0.103		−0.092
		(0.082)		(0.068)
χ^2 sig.	0.02	0.03	0.02	0.02
$\overline{R}^2$	0.09	0.09	0.10	0.11

Note: The table shows the extension to 2014Q2 of the results reported in Table 1, using both the real federal funds rate and the nominal federal funds rate to measure changes in monetary policy. The first survey used is from 1970Q4, so the sample begins with the survey in 1972Q1 because we need five survey lags to calculate the $FF1$ variable. Bold numbers indicate coefficients that are statistically significantly different from zero at the 5 percent level.

robust to the choice of proxy variable for measuring monetary policy.

Table 5 shows the results of extending the Rudebusch and Williams method from Table 3 to 2014Q2 and beginning the sample in 1970Q4. The results confirm those for the shorter sample: There is evidence that the four-quarter-ahead SPF forecasts are not efficient in the full sample that begins in 1970Q4 but not for the modern sample that begins in 1988Q1 and ends in 2014Q2. The results suggest that in the early years of the sample, the SPF forecasters were not efficient in using the information about the yield spread in forecasting output growth, but they became more efficient in doing so over time.

B. Alternative Starting Dates

Because of the evidence in the preceding section that suggests a change in the efficiency of the SPF output forecasts beginning in the 1980s, we look for additional evidence in support of that view by changing the starting dates of the samples. The idea is not to make a formal statistical test (because we will run many different tests that are not independent) but rather to look for evidence that something fundamental has changed in either the way the forecasters use data on monetary policy or the yield spread or that the predictive power of the variables

Table 5—Rudebusch–Williams Results with Yield Spread Alone, Extended Sample

Regression: $e_{t+h|t-1} = \alpha + \gamma S_{t-1} + \epsilon_{t+h|t-1}$

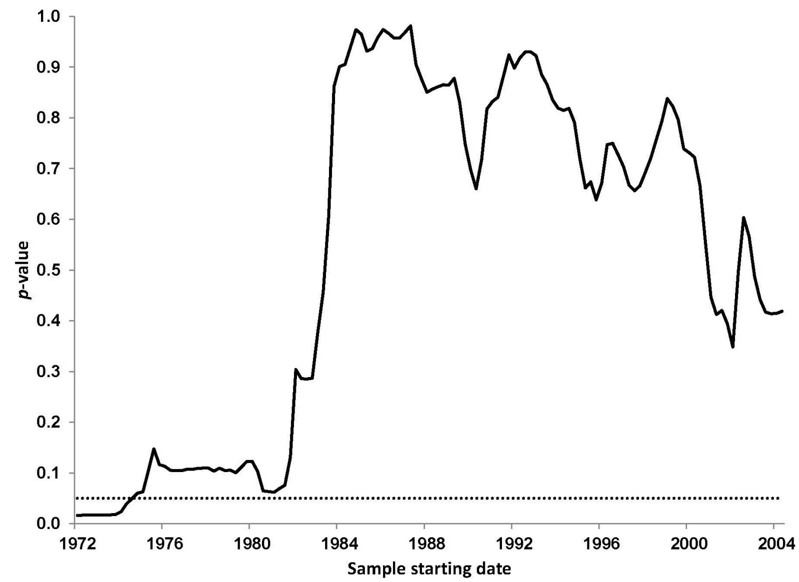
	Full Sample	Post-1987 Sample
	Current-quarter forecast	
Constant	0.15	0.36
Yield spread	0.09	−0.02
F-test (p-value)	0.11	0.16
	One-quarter-ahead forecast	
Constant	−0.52	0.00
Yield spread	0.33	0.04
F-test (p-value)	0.34	0.82
	Two-quarter-ahead forecast	
Constant	−0.92	−0.35
Yield spread	0.42	0.14
F-test (p-value)	0.38	0.83
	Three-quarter-ahead forecast	
Constant	−1.23	−0.55
Yield spread	0.48	0.16
F-test (p-value)	0.18	0.77
	Four-quarter-ahead forecast	
Constant	**−1.77**	−0.87
Yield spread	**0.72**	0.32
F-test (p-value)	0.03	0.65

Note: The table shows the rationality test used by Rudebusch and Williams (2009) but with only the yield spread used in the regression. Bold numbers indicate coefficients that are statistically significantly different from zero at the 5 percent level. The sample begins with forecasts made in 1970Q4 and ends with forecasts made in 2014Q2.

has changed over time. This method was used by Croushore (2012) to explain why researchers find differing outcomes of bias tests, depending on their choice of which survey of forecasts to examine because each different survey began at a different date. We will summarize the results by examining how the key p-values reported in Tables 4 and 5 change over time as we change the starting dates of the sample.

For the Ball and Croushore study, we examine alternative survey sample starting dates, beginning with the survey forecast in 1972Q1. Using the real federal funds rate as a measure of monetary policy, with just the $FF1$ term, which is the first column in Table 4, and changing the sample starting date from 1972Q1 to 2004Q2 (which allows us to have a 10-year sample period or longer in each case), we obtain a set of p-values from 1972Q1 to 2004Q2. The results are shown in Figure 5.

Figure 5. Alternative Sample Starting Dates, Ball–Croushore

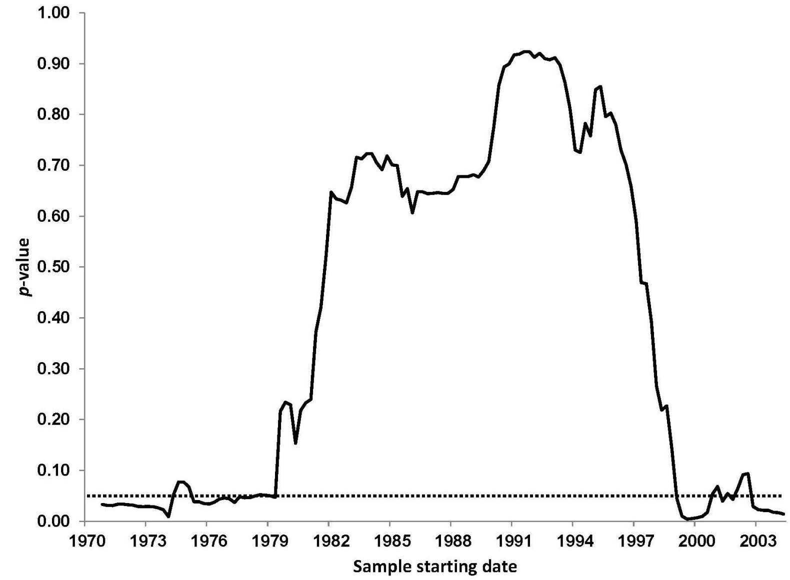


Note: The figure shows the p-values of the test for forecast efficiency, as in the first column of Table 4, but with different sample starting dates. The p-values are not overall tests because of multiple testing but show what a researcher who started the sample at the date shown on the horizontal axis would have found for a sample that ends with forecasts made in 2014Q2.

The results suggest that only samples beginning in the early 1970s feature inefficiency. Had the SPF begun later, we would find no evidence of inefficiency. In particular, in the early 1980s, the p-values jump up dramatically, suggesting a change in the behavior of monetary policy, its impact on output, or the SPF.

Doing a similar exercise for the Rudebusch and Williams study, we find mixed results, as shown in Figure 6. For starting dates early in the sample, the yield spread is significant and leads to the rejection of the test for forecast efficiency. As we start the sample in the 2000s, however, it is a significant constant term, rather than the yield spread, that leads to the rejection of efficiency. The significant constant term suggests that the forecasts were biased in that period, perhaps due to the deep recession associated with the financial crisis that began in 2008.

Figure 6. Alternative Sample Starting Dates, Rudebusch–Williams



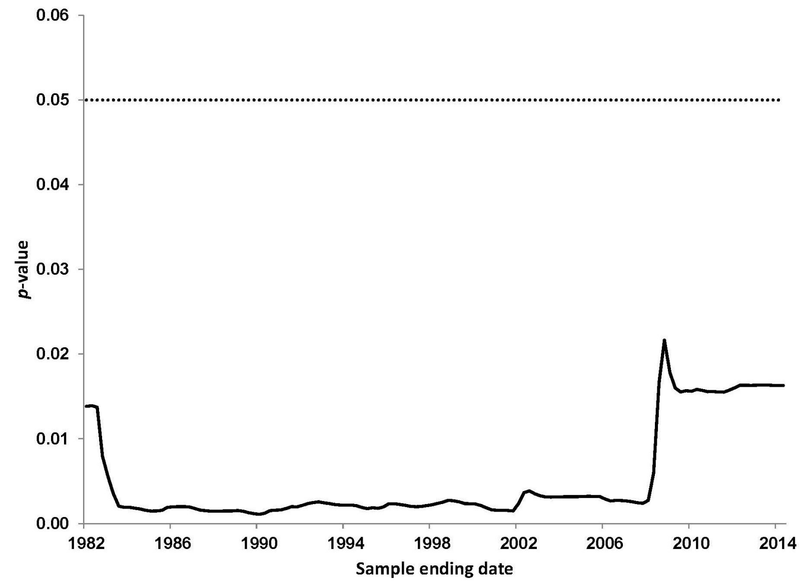
Note: The figure shows the p-values of the test for forecast efficiency, as in the first column of Table 5 for the four-quarter horizon, but with different sample starting dates, from 1970Q4 to 2004Q2. The p-values are not overall tests because of multiple testing but show what a researcher who started the sample at the date shown on the horizontal axis would have found for a sample that ends with forecasts made in 2014Q2.

C. *Alternative Ending Dates*

Given that the starting date for the evaluation of forecast efficiency matters, what if we change the end date of the sample instead? The idea of this experiment is to imagine researchers at different points in time using the SPF to evaluate forecast efficiency. If the results are the same, regardless of the ending date of the sample, we could argue that the finding of inefficiency is robust. But if the results of the test change over time, then perhaps the outcome is not robust but is special to a particular sample period, which would make it difficult to make more general conclusions from the results.

We begin in Figure 7 by rerunning the Ball and Croushore empirical work for a sample that begins in 1972Q1 and ends at various dates from 1982Q1 to 2014Q2. The results show that for almost every sample ending date, we reject the null hypothesis of forecast efficiency, so the result is quite general.

FIGURE 7. ALTERNATIVE SAMPLE ENDING DATES, BALL–CROUSHORE



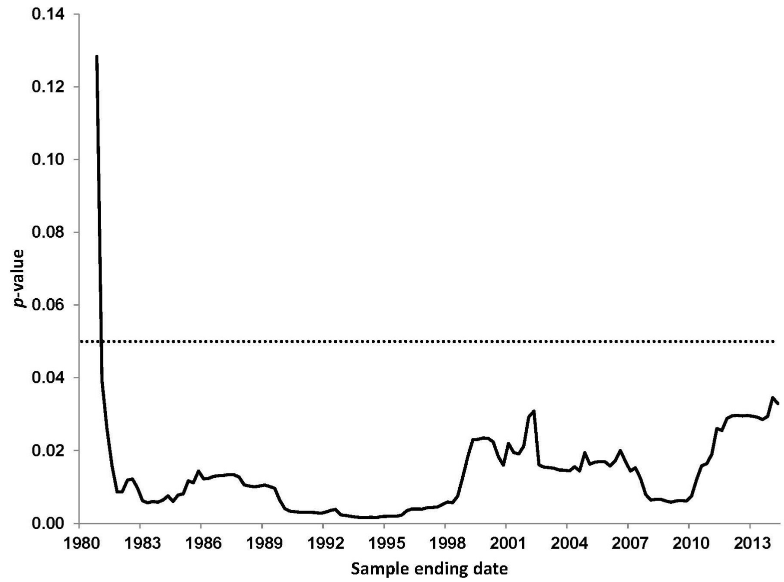
Note: The figure shows the p-values of the test for forecast efficiency, as in the first column of Table 4, but with different sample ending dates. The p-values are not overall tests because of multiple testing but show what a researcher would have found if he or she started the sample with forecasts made in 1972Q1 and ending with forecasts made at the date shown on the horizontal axis.

Engaging in the same exercise for the Rudebusch and Williams paper, we find again that the results are consistent with fairly robust conclusions against the efficiency of the SPF forecasts because for almost every sample ending date, there is evidence of inefficiency (Figure 8).

D. Alternative Actuals

Does the choice of variable being used as "actual" matter? In everything we have done so far, we have used the so-called first-final data as the actual value that is used in assessing forecast accuracy. But are the results sensitive to that choice? To investigate, we run the same analysis that we showed in Tables 4 and 5, comparing the results using first-final data (under the heading First), annual data (Annual), and pre-benchmark data (Pre-bench). The annual data are those

Figure 8. Alternative Sample Ending Dates, Rudebusch–Williams



Note: The figure shows the p-values of the test for forecast efficiency, as in the first column of Table 5, for the four-quarter horizon, but with different sample ending dates. The p-values are not overall tests because of multiple testing but show what a researcher who started the sample with forecasts made in 1970Q4 and ending with forecasts made at the date shown on the horizontal axis would have found.

coming from the first annual revision, which is usually released at the end of July of the following year, while the pre-benchmark data come from the last release of the data prior to a benchmark revision in which major methodological changes are made.

For the Ball and Croushore study, shown in Table 6, we see that the results are quite robust to the choice of actual variable. The coefficient on the $FF1$ term hardly changes at all, as is also true of the significance level and $\overline{R}^2$ statistic. So, the results of Ball and Croushore hold up very well to alternative choices of measuring the actual value of GDP growth.

The Rudebusch and Williams study is slightly more sensitive than the Ball and Croushore study to the choice of actual measure of GDP but not dramatically so. For either the annual concept of actual or the pre-benchmark concept, the results are nearly identical to the results using the first-final concept, as Table 7 shows.

Table 6—Ball–Croushore Results, Alternative Actuals

Regression: $e_t = \beta FF1_t + \epsilon_t$			
Interest rate		Real	
	First	Annual	Pre-bench
$FF1$	**−0.375**	**−0.375**	**−0.363**
	(0.156)	(0.154)	(0.170)
χ^2 sig.	0.02	0.02	0.03
$\overline{R}^2$	0.09	0.08	0.07
Interest rate		Nominal	
	First	Annual	Pre-bench
$FF1$	**−0.333**	**−0.326**	**−0.331**
	(0.141)	(0.134)	(0.148)
χ^2 sig.	0.02	0.02	0.03
$\overline{R}^2$	0.10	0.09	0.09

Note: The table shows the results of using three alternative measures of actual GDP growth: First, which is the first final (used earlier); Annual, the annual revision; and Pre-bench, the pre-benchmark revision. As in Table 4, we compare the results using both the real federal funds rate and the nominal federal funds rate to measure changes in monetary policy. Bold numbers indicate coefficients that are statistically significantly different from zero at the 5 percent level. The sample begins with forecasts made in 1972Q1 and ends with forecasts made in 2014Q2 for the First actual, 2013Q4 for the Annual actual, and 2014Q3 for the Pre-bench actual.

The only change in the significance of any result is that with the annual concept of actual, the yield spread is significantly different from zero for the two-quarter-ahead forecast. The overall significance of the regression does change slightly for some horizons, especially for the three-quarter-ahead forecast. But in general, the conclusions do not change, in that the only significant results occur for the full sample, and no variables are ever significant in the post-1987 sample.

E. Rolling Windows

Given that the results appear to depend on the exact sample period chosen, especially for the Rudebusch and Williams results, a useful technique to explore the sensitivity of the results is to examine rolling windows of forecasts to see if there are periods when the forecasts were inefficient and other periods when the forecasts were efficient. We chose to examine 10-year rolling windows of forecasts and examine the p-value for forecast efficiency in the regression equation. For the

Table 7—Rudebusch–Williams Results with Yield Spread Alone, Alternative Actuals

Regression: $e_{t+h|t-1} = \alpha + \gamma S_{t-1} + \epsilon_{t+h|t-1}$

	Full Sample			Post-1987 Sample		
	First	Annual	Pre-Bench	First	Annual	Pre-Bench
			Current-quarter forecast			
Constant	0.15	0.00	−0.04	0.36	−0.10	−0.02
Yield spread	0.09	0.18	0.18	−0.02	0.19	0.12
F-test (p-value)	0.11	0.08	0.13	0.16	0.08	0.43
			One-quarter-ahead forecast			
Constant	−0.52	−0.63	−0.71	0.00	−0.40	−0.31
Yield spread	0.33	0.38	0.41	0.04	0.19	0.12
F-test (p-value)	0.34	0.18	0.24	0.82	0.47	0.87
			Two-quarter-ahead forecast			
Constant	−0.92	−1.09	−1.15	−0.35	−0.69	−0.57
Yield spread	0.42	0.51	0.50	0.14	0.27	0.16
F-test (p-value)	0.38	0.13	0.26	0.83	0.38	0.69
			Three-quarter-ahead forecast			
Constant	−1.23	**−1.45**	**−1.46**	−0.55	−0.90	−0.76
Yield spread	0.48	**0.62**	**0.59**	0.16	0.31	0.20
F-test (p-value)	0.18	0.02	0.10	0.77	0.62	0.52
			Four-quarter-ahead forecast			
Constant	**−1.77**	**−1.90**	**−1.92**	−0.87	−1.13	−1.06
Yield spread	**0.72**	**0.81**	**0.77**	0.32	0.41	0.33
F-test (p-value)	0.03	0.00	0.02	0.65	0.52	0.47

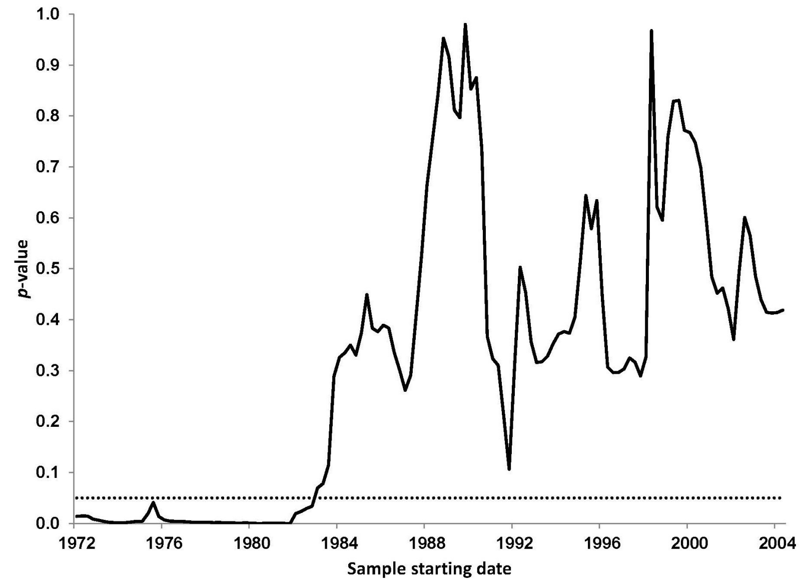
Note: The table shows the rationality test used by Rudebusch and Williams (2009) but with only the yield spread used in the regression, using three alternative measures of actual GDP growth: First, which is the first final (used earlier); Annual, the annual revision; and Pre-Bench, the pre-benchmark revision. Bold numbers indicate coefficients that are statistically significantly different from zero at the 5 percent level. The sample begins with forecasts made in 1970Q4 and ends with forecasts made in 2014Q2 for the First actual, 2013Q4 for the Annual actual, and 2014Q3 for the Pre-Bench actual.

Ball and Croushore case, we are examining the first column of Table 4 but with rolling 10-year samples. Figure 9 shows the results.

The results of the rolling-window exercise are interesting and somewhat surprising. For 10-year samples that begin before about 1982, there is evidence of inefficiency, but for samples that begin after that, there is no evidence of inefficiency for a single 10-year window. The results suggest that something has changed in the nature of the forecasts in the early 1980s.

Performing the same type of analysis with the Rudebusch and Williams study, shown in Figure 10, we find similar results. After samples that begin in the early 1980s, there are only a few 10-year windows for which the p-value falls below

Figure 9. Rolling Windows, Ball–Croushore



Note: The figure shows the p-values of the test for forecast efficiency, as in the first column of Table 4 but with rolling 10-year samples, beginning with the date shown on the horizontal axis. The p-values are not overall tests because of multiple testing but show what a researcher who started the sample at the date shown on the horizontal axis would have found for a sample that ends 10 years later.

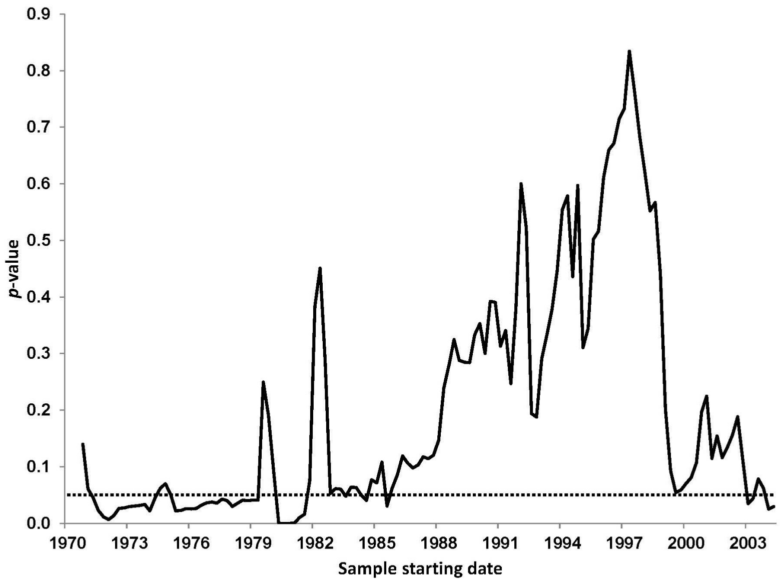
0.05. This again suggests that something about the forecasts changed in the early 1980s.

V. Did the Nature of the SPF Change in the Great Moderation?

The results of the starting-date and rolling-window analysis suggest that the SPF forecast errors before the early 1980s are the key source of the inefficiency results for both studies. The p-values jump up sharply for sample periods that begin after the early 1980s. Stock and Watson (2003) suggest that the Great Moderation began in 1984, so in this section, we test to see if our improved efficiency results are consistent with better forecasts beginning at that date.

If we rerun the empirical results for the Ball and Croushore study in Table 4 using only the SPF sample beginning in 1984Q1, we find that the inefficiency that was apparent over the full sample disappears completely, as Table 8 shows.

Figure 10. Rolling Windows, Rudebusch–Williams



Note: The figure shows the p-values of the test for forecast efficiency, as in the first column of Table 5, but with rolling 10-year samples, beginning with the date shown on the horizontal axis. The p-values are not overall tests because of multiple testing but show what a researcher who started the sample at the date shown on the horizontal axis would have found for a sample that ends 10 years later.

Comparing the results in Table 4 with those in Table 8, we see quite different results. No coefficients are statistically significantly different from zero, and the $\overline{R}^2$ statistics are zero or negative. Thus, it appears that the forecasters are using information on monetary policy efficiently since the beginning of the Great Moderation.

Performing the same exercise for the Rudebusch and Williams study, we find in Table 9 that the SPF sample beginning in 1984Q1 shows no evidence of the inefficiency that we found for the sample beginning in 1970Q4. We had found the same lack of inefficiency in the forecasts for the sample that began in 1988Q1; these results suggest that the beginning of the Great Moderation in 1984 may have had a lot to do with the decline in inefficiency found in our extension of the Rudebusch and Williams results.

Table 8—Ball–Croushore Results and Replication, Sample Beginning 1984Q1

Regression: $e_t = \beta_1 FF1_t + \beta_2 FF2_t + \epsilon_t$

Interest rate	Real		Nominal	
$FF1$	−0.015	0.009	−0.039	−0.016
	(0.122)	(0.139)	(0.108)	(0.126)
$FF2$		−0.161		−0.133
		(0.159)		(0.131)
χ^2 sig.	0.90	0.42	0.72	0.29
$\overline{R}^2$	−0.02	0.00	−0.01	0.00

Note: The table shows the results reported in Table 4 but starting the sample with forecasts made in 1984Q1. Bold numbers indicate coefficients that are statistically significantly different from zero at the 5 percent level. The sample ends with forecasts made in 2014Q2.

Table 9—Rudebusch–Williams Results with Yield Spread Alone, Sample Beginning 1984Q1

Regression: $e_{t+h|t-1} = \alpha + \gamma S_{t-1} + \epsilon_{t+h|t-1}$

	1970Q4 to 2014Q2	1984Q1 to 2014Q2
	Current-quarter forecast	
Constant	0.15	0.39
Yield spread	0.09	−0.05
F-test (p-value)	0.11	0.23
	One-quarter-ahead forecast	
Constant	−0.52	0.04
Yield spread	0.33	−0.01
F-test (p-value)	0.34	1.00
	Two-quarter-ahead forecast	
Constant	−0.92	−0.22
Yield spread	0.42	0.04
F-test (p-value)	0.38	0.88
	Three-quarter-ahead forecast	
Constant	−1.23	−0.43
Yield spread	0.48	0.09
F-test (p-value)	0.18	0.65
	Four-quarter-ahead forecast	
Constant	**−1.77**	−0.76
Yield spread	**0.72**	0.27
F-test (p-value)	0.03	0.72

Note: The table shows the rationality test from Table 5, comparing the results from starting the sample with forecasts made in 1970Q4 with the results from starting the sample with forecasts made in 1984Q1. Bold numbers indicate coefficients that are statistically significantly different from zero at the 5 percent level. The sample ends with forecasts made in 2014Q2.

VI. Forecast-Improvement Exercises

The true test of inefficiency in forecasting is to see if inefficiencies in a forecast can be improved in real time. Such forecast-improvement exercises are rare in the literature but are very convincing, as in the results of Faust, Rogers and Wright (2005). The idea is to simulate a real-time forecast, using only real-time data and without peeking at future data, to illustrate how one could use regressions that show inefficiency period by period to modify the SPF forecast to make a better forecast. It is an out-of-sample exercise that suggests how one could exploit the forecast inefficiency.

To run a forecast-improvement exercise, we take the regression results from the tests for inefficiency and apply them in real time to modify the SPF survey forecast. For example, in the Ball and Croushore study, the main regression with significant coefficients was

$$e_t = \alpha FF1_t + \epsilon_t. \tag{9}$$

Taking the estimated $\hat{\alpha}$, using it in the regression and recalling that $e_t = y_t - y_t^e$, we create, at each date t, an improved forecast y_t^f, where

$$y_t^f = y_t^e + \hat{\alpha} FF1_t. \tag{10}$$

Thus, we are using the regression to modify the original SPF forecast to try to make a new forecast that is closer to the realized value based on the past relationship between forecast errors and the most recently observed change in monetary policy. We need a number of years of SPF forecasts before we can reasonably engage in this exercise so that the sampling error is not too large. So, we will start the forecast-improvement exercise in 1983Q2, using the forecast errors (based on first-final actuals) from 1972Q1 to 1982Q1 to estimate Equation (9), and then use the estimated coefficient to improve on the four-quarter-ahead SPF forecast made in 1983Q2. Then, we step forward one quarter, use the forecast-error data from surveys from 1972Q1 to 1982Q2, re-estimate Equation (9), and then use the estimated coefficient to improve on the four-quarter-ahead SPF forecast made in 1983Q3. We keep repeating this process through to 2014Q2. At the end of the process, we compare the root-mean-squared error (RMSE) for these improved forecasts with the RMSE for the original SPF forecasts to see if indeed the forecasts have been improved upon or not. We do the same type of exercise using five- and 10-year rolling windows to see if that works better.

The results of the forecast-improvement exercises for the Ball and Croushore analysis are shown in Table 10. The results show that attempting to improve on the SPF survey results always makes the RMSE higher, but the damage from trying to do so is never so large as to be statistically significant.

We do the same exercise for the Rudebusch and Williams study. In this case, we base our analysis on the equation

$$e_{t+h|t-1} = \alpha + \gamma S_{t-1} + \epsilon_{t+h|t-1}. \tag{11}$$

We try to improve the SPF forecast in an analogous fashion to that in the Ball and Croushore study:

$$y^f_{t+h|t-1} = y^e_{t+h|t-1} + \hat{\alpha} + \hat{\gamma} S_{t-1}. \tag{12}$$

The results of the forecast-improvement exercise are shown in Table 11. We find RMSE ratios always greater than 1, which means the attempt to improve on the survey makes the forecasts worse. The results are statistically significantly worse for several different horizons and for both expanding samples and rolling

Table 10—Ball–Croushore Forecast-Improvement Exercise

Interest rate	Real	Nominal
Expanding window		
RMSE of survey	1.587	1.587
RMSE of improvement to survey	1.649	1.614
RMSE ratio	1.039	1.017
p-value	0.35	0.68
Ten-year rolling window		
RMSE of survey	1.587	1.587
RMSE of improvement to survey	1.608	1.632
RMSE ratio	1.013	1.028
p-value	0.45	0.40
Five-year rolling window		
RMSE of survey	1.530	1.530
RMSE of improvement to survey	1.629	1.962
RMSE ratio	1.065	1.282
p-value	0.20	0.34

Note: The table shows results of the forecast-improvement exercise. The RMSE of both the survey and the improvement to the survey are reported in the first two rows of the results. The RMSE ratio is the RMSE of the improvement to the survey divided by the RMSE of the survey, so a number less than 1 indicates that the attempt to improve the survey was successful, while a number greater than 1 indicates that the attempt to improve upon the survey failed. The p-value is the result of the Harvey, Leybourne and Newbold (1997) variation of the Diebold and Mariano (1995) test for significant forecast differences. The sample period is 1972Q1 to 2014Q2.

samples.

VII. Ending the Sample Before the Great Recession

We might expect that our empirical results could be thrown off for the period after 2007Q4 when short-term interest rates came close to their zero lower bound. So, we repeat several of the main results in which we truncate the sample in 2007Q4 to ensure that the results are robust and not affected by the zero lower bound.

Table 12 shows that ending the Ball and Croushore study in 2007Q4 has some quantitative impact on the results, but the broad conclusions remain unchanged. For the sample beginning in 1972Q1, the quantitative results are a bit stronger, with slightly larger (in absolute value) coefficients on $FF1$ and higher values of

Table 11—Rudebusch–Williams Forecast-Improvement Exercise

Horizon	0Q	1Q	2Q	3Q	4Q
Expanding window					
RMSE of survey	1.793	2.095	2.198	2.358	2.420
RMSE of improvement to survey	2.240	2.285	2.516	2.557	2.512
RMSE ratio	1.250	1.091	1.144	1.084	1.038
p-value	0.06	0.27	0.03	0.23	0.58
Ten-year rolling window					
RMSE of survey	1.793	2.095	2.198	2.358	2.420
RMSE of improvement to survey	2.263	2.318	2.471	2.563	2.601
RMSE ratio	1.262	1.107	1.124	1.087	1.074
p-value	0.04	0.25	0.01	0.01	0.03
Five-Year Rolling Window					
RMSE of survey	1.726	1.983	2.113	2.281	2.340
RMSE of improvement to survey	2.143	2.257	2.643	2.656	2.563
RMSE ratio	1.242	1.138	1.251	1.164	1.095
p-value	0.05	0.17	0.03	0.09	0.32

Note: The table shows results of the forecast-improvement exercise. The RMSE of both the survey and the improvement to the survey are reported in the first two rows of the results. The RMSE ratio is the RMSE of the improvement to the survey divided by the RMSE of the survey, so a number less than 1 indicates that the attempt to improve the survey was successful, while a number greater than 1 indicates that the attempt to improve upon the survey failed. The p-value is the result of the Harvey, Leybourne and Newbold (1997) variation of the Diebold and Mariano (1995) test for significant forecast differences. The sample period is 1970Q4 to 2014Q2.

$\overline{R}^2$. So, it may be that the data from 2008Q1 to 2014Q2 weakened the relationship between changes in real interest rates and output forecast errors. Similarly, for the sample that begins in 1984Q1, truncating the sample at 2007Q4 also leads to slightly stronger results, again with slightly larger (in absolute value) coefficients on $FF1$ and higher values of $\overline{R}^2$. But none of the coefficients are statistically significantly different from zero as we found in Table 8.

Table 13 shows that for the Rudebusch–Williams case, results are quite similar whether the sample ends in 2014Q2 or 2007Q4. The broad conclusion remains that the yield spread was significantly related to forecast errors for samples with forecasts beginning in 1970Q4 but not for samples with forecasts beginning in 1984Q1. This result confirms the robustness of the conclusion that the yield spread no longer has predictive power since the beginning of the Great Modera-

Table 12—Ball–Croushore Results and Replication, Great Recession Effect

Regression: $e_t = \beta_1 FF1_t + \beta_2 FF2_t + \epsilon_t$

Interest rate	Real		Nominal	
	Sample 1972Q1 to 2007Q4			
$FF1$	**−0.449**	**−0.443**	**−0.385**	**−0.374**
	(0.148)	(0.153)	(0.138)	(0.142)
$FF2$		−0.095		−0.094
		(0.078)		(0.065)
χ^2 sig.	<0.01	0.01	0.01	0.01
$\overline{R}^2$	0.16	0.17	0.18	0.18
	Sample 1984Q1 to 2007Q4			
$FF1$	−0.151	−0.133	−0.140	−0.123
	(0.089)	(0.099)	(0.090)	(0.101)
$FF2$		−0.133		−0.130
		(0.147)		(0.118)
χ^2 sig.	0.09	0.10	0.12	0.05
$\overline{R}^2$	0.02	0.03	0.03	0.05

Note: The table shows what happens to the results reported in Tables 4 and 8 when we end the sample at 2007Q4, using both the real federal funds rate and the nominal federal funds rate to measure changes in monetary policy. Bold numbers indicate coefficients that are statistically significantly different from zero at the 5 percent level.

tion.

VIII. Summary and Conclusions

There is some evidence in both the Ball and Croushore study and the Rudebusch and Williams study that the SPF forecasts of GDP growth are not efficient with respect to the yield spread and to measures of monetary policy. But the evidence suggests that the main inefficiencies arose in the early years of the survey and not since the Great Moderation, which began around 1984. In addition, the forecast-improvement exercises suggest that even the inefficiency that existed then was not large enough to be exploitable in real time.

The results may not be too surprising, given the incentive of forecasters to exploit observable patterns in the data. Perhaps as economists discovered the

Table 13—Rudebusch–Williams Results, Great Recession Effect

Regression: $e_{t+h|t-1} = \alpha + \gamma S_{t-1} + \epsilon_{t+h|t-1}$

	1970Q4 to 2014Q2	1970Q4 to 2007Q4	1984Q1 to 2014Q2	1984Q1 to 2007Q4
	Current-quarter forecast			
Constant	0.15	0.13	0.39	0.36
Yield spread	0.09	0.17	−0.05	0.04
F-test (p-value)	0.11	0.03	0.23	0.10
	One-quarter-ahead forecast			
Constant	−0.52	−0.52	0.04	0.08
Yield spread	0.33	0.42	−0.01	0.05
F-test (p-value)	0.34	0.17	1.00	0.75
	Two-quarter-ahead forecast			
Constant	−0.92	−0.83	−0.22	−0.01
Yield spread	0.42	0.47	0.04	0.04
F-test (p-value)	0.38	0.23	0.88	0.94
	Three-quarter-ahead forecast			
Constant	−1.23	−1.12	−0.43	−0.15
Yield spread	0.48	**0.53**	0.09	0.07
F-test (p-value)	0.18	0.09	0.65	0.96
	Four-quarter-ahead forecast			
Constant	**−1.77**	**−1.72**	−0.76	−0.59
Yield spread	**0.72**	**0.78**	0.27	0.29
F-test (p-value)	0.03	0.01	0.72	0.44

Note: The table extends Table 9, comparing the results from ending the sample with forecasts made in 2014Q2 with the results from ending the sample with forecasts made in 2007Q4. Bold numbers indicate coefficients that are statistically significantly different from zero at the 5 percent level.

forecast inefficiency in the early 1980s during the rational-expectations revolution, forecasters modified their forecasting methods and became more efficient. Or perhaps the Great Moderation itself made forecasting easier, and the apparent inefficiencies disappeared. Whatever the cause, the results of inefficiency in the SPF forecasts found by Ball and Croushore, as well as by Rudebusch and Williams, are no longer observable, even when we restrict the sample to the period prior to the Great Recession of 2007–2009.

REFERENCES

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

Croushore, Dean. 1993. "Introducing: The Survey of Professional Forecasters." *Business Review*, Federal Reserve Bank of Philadelphia, 3–13.

Croushore, Dean. 2010. "Philadelphia Fed Forecasting Surveys: Their Value for Research." *Business Review*, Federal Reserve Bank of Philadelphia, 1–11.

Croushore, Dean. 2012. "Forecast Bias in Two Dimensions." *Working Paper*, Federal Reserve Bank of Philadelphia, 12-9.

Croushore, Dean, and Katherine Marsten. 2015. "Reassessing the Relative Power of the Yield Spread in Forecasting Recessions." *Journal of Applied Econometrics*, DOI: 10.1002/jae.2485.

Croushore, Dean, and Tom Stark. 2001. "A Real-Time Data Set for Macroeconomists." *Journal of Econometrics*, 105: 111–130.

Diebold, Francis X., and Roberto S. Mariano. 1995. "Comparing Predictive Accuracy." *Journal of Business and Economic Statistics*, 13: 253–263.

Faust, Jon, John H. Rogers, and Jonathan H. Wright. 2005. "News and Noise in G-7 GDP Announcements." *Journal of Money, Credit and Banking*, 37: 403–419.

Gneiting, Tilmann. 2011. "Making and Evaluating Point Forecasts." *Journal of the American Statisticial Association*, 106(494): 746–762.

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

Holdren, Alyssa E. 2014. "Gross Domestic Product and Gross Domestic Income: Revisions and Source Data." *Survey of Current Business*, 94: 1–11.

Pesaran, M. Hashem, and Martin Weale. 2006. "Survey Expectations." In *Handbook of Economic Forecasting.* Vol. 1, ed. Graham Elliott, Clive W.J. Granger and Allan Timmermann, 715–776. Elsevier.

Rudebusch, Glenn D., and John C. Williams. 2009. "Forecasting Recessions: The Puzzle of the Enduring Power of the Yield Curve." *Journal of Business and Economic Statistics*, 27(4): 492–503.

Stock, James H., and Mark W. Watson. 2003. "Has the Business Cycle Changed? Evidence and Explanations." *Symposium on Monetary Policy and Uncertainty: Adapting to a Changing Economy*, Federal Reserve Bank of Kansas City, 9–56.

Zarnowitz, Victor, and Phillip Braun. 1993. "Twenty-two Years of the NBER-ASA Quarterly Economic Outlook Surveys: Aspects and Comparisons of Forecasting Performance." In *Business Cycles, Indicators and Forecasting.* , ed. James H. Stock and Mark W. Watson, 11–94. University of Chicago Press.