WORKING PAPERS
RESEARCH DEPARTMENT

WORKING PAPER NO. 14-5
THE CONTINUING POWER OF THE YIELD SPREAD IN FORECASTING RECESSIONS

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

Katherine Marsten
University of Richmond

February 13, 2014

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

The Continuing Power of the Yield Spread in Forecasting Recessions

By Dean Croushore and Katherine Marsten*

February 13, 2014

In this paper, we replicate the main results of Rudebusch and Williams (2009), who show that the use of the yield spread in a probit model can predict recessions better than the Survey of Professional Forecasters. We investigate the robustness of their results in several ways: extending the sample to include the 2007–09 recession, changing the starting date of the sample, changing the ending date of the sample, using rolling windows of data instead of just an expanding sample, and using alternative measures of the "actual" value of real output. Our results show that the Rudebusch-Williams findings are robust in all dimensions. *Keywords: real-time data, recession forecasts, yield spread*

Economists' ability to predict recessions is limited because recessions are often caused by shocks that cannot be anticipated. For example, the 1990–91 recession seems to have been caused by the sudden rise in oil prices at the start of the Gulf War. Nonetheless, because the impact of recessions is so substantial, economists devote much time to the effort of predicting them. The strong correlation between the yield spread (the interest rate on a long-term bond minus the interest rate on a short-term bond) and recessions has led economists to use the spread as a

* Croushore: University of Richmond, Robins School of Business, dcrousho@richmond.edu. Marsten: University of Richmond, katherine.marsten@richmond.edu. We thank the University of Richmond for summer research support. Thanks to Kajal Lahiri for sharing his data and programs. 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/.

variable that can potentially be used to forecast recessions. For that reason, a recent paper by Rudebusch and Williams (2009) is particularly interesting because it shows that the yield spread provides better forecasts of recessions than does the Survey of Professional Forecasters (SPF), suggesting that professional forecasters are not using the yield spread optimally in their forecasting models.

Our analysis thus digs more deeply into the results of Rudebusch and Williams, replicating them and examining their robustness in several dimensions to see how well they hold up to increased scrutiny. Their results are 1) a probit model that forecasts recessions based on the yield spread between the interest rate on 10-year Treasury bonds and three-month Treasury bills is superior to the forecasts of the SPF for longer-horizon (three or four quarters ahead) forecasts; and 2) the results hold up even when the sample starts after the Great Moderation begins. We examine the sensitivity of their results to the starting and ending dates of their analysis because previous research has shown that forecast efficiency is often sensitive to that choice. We also examine whether their choice of which data vintage to use as "actual" in calculating forecast errors matters for their results.

In this paper, we begin by discussing the extant literature on the use of the yield curve in forecasting recessions, including the details of Rudebusch and Williams and the importance of their results. Next, we describe the data on the SPF, the real-time data set used to determine the "actual" values of variables being forecast, and the determination of recession dates. We then replicate the main results of Rudebusch and Williams to show that our methods are not different from theirs. Our robustness exercises include extending the sample to include the 2007–09 recession, changing the starting date of the sample, changing the ending date of the sample, using rolling windows of data instead of just an expanding sample, and using alternative measures of the "actual" value of real output.

Our main conclusions are that, using the methods of Rudebusch and Williams, we can confirm their results on the power of the yield spread to predict recessions. Their general results hold up across many alternative sample periods and alternative versions of the data, including real-time data.

I. Research on Forecasting Recessions

Researchers have become more pessimistic over time about economists' ability to forecast recessions. Early on, Estrella and Hardouvelis (1991) showed that the yield spread was useful in predicting economic activity and recessions in the United States, and Bernard and Gerlach (1996) looked at forecasting recessions with the yield spread in many other countries. However, Stock and Watson (2003) highlight the fact that many studies (including the two cited above) that found predictive ability of the yield curve did not perform their analysis in real time, but used only in-sample fit. Out-of-sample analysis of the yield spread's usefulness in forecasting recessions shows little forecasting power for the yield spread, especially since 1985. In fact, Stock and Watson's own experimental leading index[1] that used the yield spread failed to predict the 1990–91 recession. Further, Diebold and Rudebusch (1991) suggest that the index of leading indicators neither leads nor indicates in real time, as opposed to its forecasting ability after the fact when it is created to work in-sample.

Rudebusch and Williams (2009) use a probit model based on the yield spread (the interest rate on 10-year U.S. Treasury bonds minus the interest rate on three-month U.S. Treasury bills) to forecast declines in real output. They compare the forecasting ability of that yield-spread probit model to the mean forecast of the probability of a decline in real output made by the participants in the Survey of Professional Forecasters. The Rudebusch-Williams analysis is based on declines in real output as measured in the first-final release of the NIPA data; that is, the real output growth rate as reported at the end of the third month following the end of the quarter. First-final data allow for some data revisions to occur, but avoid the problems caused by redefinitions of variables that occur in benchmark revisions.

[1]See Stock and Watson (1989).

Rudebusch and Williams find that, although the SPF forecasts of declines in real output outperform the yield-spread probit forecasts at very short horizons, especially during the current quarter, the yield-spread probit forecasts are statistically significantly better at longer horizons, especially three and four quarters ahead. The results are robust to using data samples that begin in 1968 or in 1987, so the forecasting ability of the yield-spread probit model does not seem to have deteriorated over time.

Given the questionable out-of-sample forecasting power of the term spread for predicting recessions found by the earlier literature, the finding by Rudebusch and Williams that the yield-spread probit model could outperform the Survey of Professional Forecasters is quite surprising. Thus, a more thorough investigation of the yield spread's out-of-sample forecasting ability seems warranted.

II. Data

The Survey of Professional Forecasters is the most well-known survey of U.S. macroeconomic forecasts and is widely used by researchers and policy analysts. The survey was begun in 1968 as a joint effort by the American Statistical Association and the National Bureau of Economic Research.[2] In 1990, those institutions gave up the survey and it was taken over by the Federal Reserve Bank of Philadelphia.[3] The participants in the survey include many practicing forecasters from Wall Street companies, consulting firms, and academia. They are asked to complete a survey on their forecasts once each quarter, immediately following the release of the NIPA data for the preceding quarter. The survey asks for the participants' quarterly-point forecasts at horizons up to four quarters ahead, as well as calendar-year forecasts for the current and following year, for 18 different macroeconomic variables, including real output and all of its components, CPI inflation, interest rates, and other broad macroeconomic variables. In addition to

[2]See Zarnowitz and Braun (1993) for a history of the early years of the survey.
[3]See Croushore (1993) for a description of the survey after the Philadelphia Fed took it over.

these point forecasts, the survey also asks participants about the distribution of their forecasts, posing questions about real output growth and inflation and the uncertainty the forecasters have about their growth rates.

The key recession-related question used by Rudebusch and Williams asks the SPF participants to report the probability that real output will decline in the current quarter and in each of the following four quarters. The SPF panelists turn over frequently, so evaluation of individuals' forecasts is problematic; instead, Rudebusch and Williams look at the mean probability forecasts across all the participants. Because the forecasts describe the probability of a decline in just one quarter, they are not strictly recession forecasts, though they are often treated that way.[4] For that reason, the outcomes for evaluating the forecasts are not recession dates, as determined by the National Bureau of Economic Research Business-Cycle Dating Committee, but rather they are the quarters in which real output declines.

Determining the quarters in which real output declines, however, is not clear-cut because of the presence of data revisions. For that reason, Rudebusch and Williams base their analysis on the so-called first-final data release, which is the value for real output released at the end of the third month following the end of the quarter in question; for example, the first-final release for the first quarter of the year is made at the end of June. To check for robustness, Rudebusch and Williams also look at the initial release of the data. However, there are other alternative data choices that may be of interest, so we will use the Real-Time Data Set for Macroeconomists, described by Croushore and Stark (2001). For example, Croushore (2012) finds a substantial amount of information in the annual revision of the NIPA data, which suggests the use of the annual revision as a standard, while Zarnowitz (1985) recommends the use of the last vintage of the data before a benchmark revision occurs. Using that vintage, the data

[4]If we ignore transition quarters (that is, quarters in which recessions begin or end), then based on data of January 2014, there have been 144 quarters of expansion based on NBER recession dates from 1968:Q4 to 2013:Q3 and 22 quarters of recession. In only three of the 144 expansion quarters did real GDP decline. But in nine of the 22 recession quarters, real GDP increased.

include the most information possible under a given methodology for computing real output.

The Real-Time Data Set for Macroeconomists (RTDSM) consists of vintages of data series for real output (GNP before 1992, GDP since 1992) and other major macroeconomic variables. Each vintage records the entire time series available to an observer at a point in time. For example, the first SPF forecast is made in early November 1968; the RTDSM contains the data that were available to the public at that time, called real-time data. In particular, the RTDSM shows, for each month from November 1965 to the present, the values of the data in the middle of each month. Since most data series are released or revised once a month, the RTDSM thus contains the entire history of the data for each variable for any observation date, with any irregular observations described in the database documentation.

The real-time data are crucial for the analysis of forecasts because the researcher must know the data available to the forecasters in the SPF if the forecasts are to be analyzed correctly. Some studies in the forecast-evaluation literature have assumed that data revisions are small and random, so they have not used real-time data. But the results in the real-time literature, described by Croushore (2011), suggest that data revisions are large and systematic. Thus, forecast evaluation research must use real-time data to be accurate.

Rudebusch and Williams begin by calculating three alternative sets of actual values for real output. Their baseline analysis is based on first-final data, described above, and they check the robustness of their results to two alternatives: 1) the initial release (also called the advance release) of the data; and 2) the final, revised data as of February 2007.[5] They use those three alternative data sets to determine the quarters in which real output declines, which they call $R1$ recession dates. These dates do not match the NBER official dates of recessions, but they are appropriate for the question about a decline in real output from the SPF.

[5] Although the Rudebusch-Williams paper says it uses the February 2007 vintage of the data, evidently in revising the paper the authors extended the data a bit further because they include the first quarter of 2007, for which data were not available until later in the year. But none of their results are affected by this difference, as we have verified.

For calculating the yield spread, which is the main variable on which Rudebusch and Williams focus, they use the interest rate on 10-year U.S. government Treasury bonds minus the interest rate on 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.

III. Replication Results

The main result in Rudebusch and Williams is a comparison of the SPF probability forecasts of a decline in real output with a probit model based on the yield spread of the probability of a decline in real output. Their basic yield-spread model is

$$Pr[R1_{t+h} = 1 | I_t] = N[\alpha + \beta S_{t-1}]. \tag{1}$$

In this equation, we think of a forecaster standing at date t, using information available at time t, I_t, which includes data on the yield spread from the previous quarter, S_{t-1}. That is, the probit model uses just the spread from the quarter before the quarter in which the SPF forecast is made and the history of past recession dates to estimate the probability of a decline in real output h quarters in the future. The variable $R1$ takes the value 1 in quarters when real output declines; it takes the value 0 when real output increases in the quarter. The forecast horizon h takes values from 0 to 4, representing a current-quarter forecast ($h = 0$), a one-quarter-ahead forecast ($h = 1$), and so on, up to a four-quarter-ahead forecast ($h = 4$). For example, for comparison with the SPF forecasts at the date of the first SPF survey, taken in the fourth quarter of 1968, the probit model in Equation (1) with $h = 1$ is estimated on the $R1$ data from 1955:Q1 to 1968:Q3 and the yield spread from 1968:Q3. The model is then used to forecast the probability of a decline in real output in 1968:Q4. Then, the model is reestimated using the same data but with $h = 2$ in Equation (1) and used to forecast the probability of a decline in real output in 1969:Q1. A similar procedure is followed to obtain forecasts for 1969:Q2, 1969:Q3, and 1969:Q4. Then, the process is repeated by advancing one quarter at a time, adding to the information set on $R1$ and the yield spread, until an entire time series of forecasts has been generated, corresponding to the SPF forecasts made at each date from 1968:Q4 to 2007:Q1 for each of the five forecast horizons.

To evaluate the probability forecasts, Rudebusch and Williams use three alternative metrics: mean-absolute error (MAE), root-mean-squared error ($RMSE$), and log probability score (LPS). They are defined as

$$MAE(h) = \frac{1}{T}\sum_{t=1}^{T} |F_{t+h|t} - R1_{t+h}|, \tag{2}$$

$$RMSE(h) = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(F_{t+h|t} - R1_{t+h})^2}, \tag{3}$$

$$LPS(h) = -\frac{1}{T}[(1 - R1_{t+h})ln(1 - F_{t+h|t}) + R1_{t+h}ln(F_{t+h|t})]. \tag{4}$$

In these equations, there are T observations, $R1_{t+h}$ is defined as before with a value of 1 in quarters in which real output declines and 0 in quarters in which real output increases, and $F_{t+h|t}$ is the forecast made at date t for the probability of a decline in real output at date $t + h$. The term F is either the forecast made by the SPF or the probit model based on the yield spread.

We calculate each of these measures for both the SPF and the yield-spread probit forecast. We compare our replication results to those of Rudebusch and Williams in Table 1. In the table, the rows that begin with RW are those reported in Rudebusch and Williams; the other rows are our replication results. We report both a sample beginning in 1968:Q4 and another beginning in 1987:Q1, as did Rudebusch and Williams, because of concern that the forecasting power of the yield spread changed in the 1980s. In the table, numbers in bold show the lowest value for each pair, comparing the SPF reported value versus the yield-spread probit value, for each pair. A dagger indicates a statistically significant difference at the 5 percent level in the MAE, RMSE, or LPS, based on a Diebold-Mariano test with HAC standard errors.

As Table 1 shows, the results of Rudebusch and Williams are generally confirmed. Our numbers differ only slightly from theirs, showing significant results in most of the same places. Thus, broadly speaking, our results confirm their main overall result: that the yield spread provides useful forecasts of declines in real output for forecasts that are three or four quarters ahead. We also confirm that there are not many differences between the results for the sample that begins in 1968 and the one that begins in 1987.

Our results are similar to the support for Rudebusch and Williams's results found in Lahiri, Monokroussos, and Zhao (2013). In that paper, the authors rerun the Rudebusch-Williams model, including a dynamic factor as well. The insignificance of the dynamic factor suggests that the SPF forecasters incorporate much macroeconomic information, including all of those variables in the dynamic factor model, but not the yield spread.

IV. Robustness Exercises

A. Extending the Sample

Because the sample of Rudebusch and Williams ended in 2007, we can extend it considerably by examining data through the first quarter of 2013. This might

Table 1
Evaluation of Real-time Probability Forecasts: First-Final Data

Probability Forecast	Full Sample MAE	Full Sample RMSE	Full Sample LPS	Post-1987 Sample MAE	Post-1987 Sample RMSE	Post-1987 Sample LPS
	Current-quarter $R1$ recession prediction					
RW SPF reported	**0.168**	**0.261**	**0.239**	**0.129**	**0.193**	**0.160**
RW yield spread	0.206	0.313	0.350	0.165	0.251	0.241
SPF reported	**0.162†**	**0.249†**	**0.218†**	**0.129**	**0.193**	**0.160**
Yield spread	0.205	0.307	0.331	0.168	0.252	0.247
	One-quarter-ahead $R1$ recession prediction					
RW SPF reported	0.212	**0.291**	**0.290**	0.154	**0.196**	**0.181**
RW yield spread	**0.186**	0.296	0.291	**0.150**	0.238	0.208
SPF reported	0.213	0.293	0.296	0.162	**0.219**	**0.202**
Yield spread	**0.175†**	**0.287**	**0.266**	**0.146**	0.246	0.217
	Two-quarter-ahead $R1$ recession prediction					
RW SPF reported	0.234	0.313	0.334	0.173	**0.214**	0.210
RW yield spread	**0.195†**	**0.303**	**0.332**	**0.151**	0.222	**0.197**
SPF reported	0.245	0.331	0.372	0.192	0.257	0.261
Yield spread	**0.181†**	**0.293**	**0.280†**	**0.139†**	**0.227†**	**0.193†**
	Three-quarter-ahead $R1$ recession prediction					
RW SPF reported	0.248	0.330	0.372	0.194	0.244	0.250
RW yield spread	**0.201†**	**0.307†**	**0.319†**	**0.144†**	**0.218**	**0.187†**
SPF reported	0.255	0.341	0.401	0.201	0.265	0.287
Yield spread	**0.197†**	**0.300†**	**0.296†**	**0.142†**	**0.222†**	**0.189†**
	Four-quarter-ahead $R1$ recession prediction					
RW SPF reported	0.258	0.342	0.400	0.205	0.259	0.272
RW yield spread	**0.206**	**0.311†**	**0.321†**	**0.147†**	**0.220**	**0.192**
SPF reported	0.266	0.352	0.426	0.204	0.267	0.291
Yield spread	**0.200†**	**0.301†**	**0.300†**	**0.139†**	**0.221†**	**0.185†**

Note: Bold numbers show the lowest value for each pair, comparing SPF reported value versus yield spread value. Daggers indicate statistically significant differences at the 5 percent level (Diebold-Mariano test with HAC standard errors). The measures MAE, RMSE, and LPS are defined in equations (2), (3), and (4). Rows labeled RW report values in Rudebusch and Williams (2009). The full sample is 1968:Q4 to 2007:Q1; the post-1987 sample is 1987:Q1 to 2007:Q1.

be useful because their sample ended just before the Great Recession of 2007–09, so we get an additional observation of a recession period. When we extend the sample in that way, we find slightly better results for the SPF, as shown in Table 2. Shaded cells in the table show cases in which there is either a change in statistical significance of the results compared with the results in Table 1, or a change in the ranking of the SPF reported versus the yield-spread forecast. Comparing Table 2 to Table 1, we see that, for shorter horizons, more cells show the SPF being statistically significantly better, and there are reversals for which, in the longer sample, the SPF is superior compared with the earlier sample. However, the conclusion that the yield spread produces better forecasts than the SPF at long horizons is confirmed in the longer sample. And the SPF forecasts are clearly better than the yield-spread forecasts only for the current quarter.

B. Alternative Starting Dates

Croushore (2010) shows that the results of forecast-bias tests for point forecasts of inflation are fragile, in that the finding of bias in the SPF is quite sensitive to the starting date of the sample. We can investigate the same issue for the probability forecasts of a decline in real output. The question is, do the results change when we change the starting date of the sample? We look at four-quarter-ahead forecasts with sample starting dates from 1968:Q4 to 1998:Q4 to see how the results change, where we allow the sample to end in 2012:Q1 (so that the last forecast we evaluate is the one that forecasters made in 2012:Q1 for the period 2013:Q1). Figure 1 shows the p-value of the significance of the difference between the yield-spread probit forecast and the SPF, based on $RMSE$, and the relative $RMSE$, defined as the $RMSE$ for the SPF forecast divided by the $RMSE$ for the yield-spread probit forecast. A relative $RMSE$ greater than 1 indicates that the yield-spread probit forecast has a lower $RMSE$.

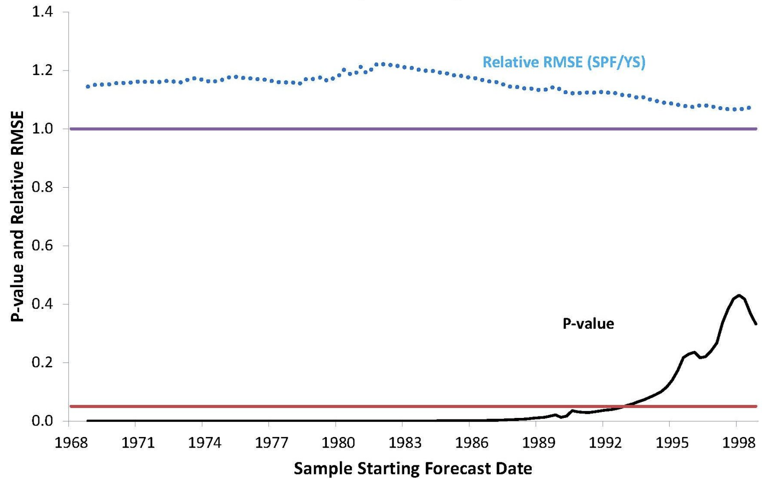
The results indicate that the main advantage to using the yield curve to forecast a decline in real GDP comes from the early years of the sample period. In Figure

Table 2
Evaluation of Real-time Probability Forecasts:
First-Final Data Sample Ending in 2012

Probability Forecast	Full Sample MAE	Full Sample RMSE	Full Sample LPS	Post-1987 Sample MAE	Post-1987 Sample RMSE	Post-1987 Sample LPS
	Current-quarter $R1$ recession prediction					
SPF reported	**0.165†**	**0.247†**	**0.218†**	**0.141†**	**0.203†**	**0.174†**
Yield spread	0.211	0.322	0.364	0.187	0.295	0.324
	One-quarter-ahead $R1$ recession prediction					
SPF reported	0.215	**0.292**	**0.296**	0.179	**0.236**	**0.226**
Yield spread	**0.186**	0.309	0.328	**0.171**	0.299	0.338
	Two-quarter-ahead $R1$ recession prediction					
SPF reported	0.252	0.337	0.381	0.217	0.289	0.305
Yield spread	**0.190†**	**0.312**	**0.325**	**0.165†**	**0.284**	**0.294**
	Three-quarter-ahead $R1$ recession prediction					
SPF reported	0.261	0.349	0.414	0.226	0.303	0.340
Yield spread	**0.202†**	**0.314†**	**0.323†**	**0.165†**	**0.274†**	**0.262†**
	Four-quarter-ahead $R1$ recession prediction					
SPF reported	0.270	0.357	0.436	0.226	0.303	0.343
Yield spread	**0.203†**	**0.312†**	**0.314†**	**0.159†**	**0.265†**	**0.239†**

Note: As in Table 1, bold numbers show the lowest value for each pair, comparing SPF reported value versus yield spread value. Daggers indicate statistically significant differences at the 5 percent level (Diebold-Mariano test with HAC standard errors). Shaded areas show changes from Table 1 in statistical significance or in which forecast (SPF reported or yield spread) has the lower value in each pair. The full sample is 1968:Q4 to 2013:Q1; the post-1987 sample is 1987:Q1 to 2013:Q1.

Figure 1
P-Values and Relative RMSEs
for Alternative Sample Starting Forecast Dates



Note: The p-values in the bottom part of the figure show the significance of the difference between the yield spread (YS) probit forecast and the SPF forecast, based on $RMSE$. The top part of the figure shows the relative $RMSE$, defined as the $RMSE$ for the SPF forecast divided by the $RMSE$ for the yield-spread probit forecast. A relative $RMSE$ greater than 1 indicates that the yield-spread probit forecast has a lower $RMSE$.

1, we observe p-values of less than 0.05, indicating that the yield-spread probit forecast is statistically significantly better than the SPF forecast, for samples that begin with the SPF surveys of 1968:Q4 through 1992:Q4. For samples that begin with the SPF surveys of 1993:Q1 to 1998:Q4, the yield-spread probit model has a lower $RMSE$ than the SPF forecast but the p-value is above 0.05, so the difference is no longer statistically significant. This result is consistent with the observation that forecasting models based on the yield spread, such as in Stock and Watson (1993), performed poorly in the 1990–91 and 2001 recessions.[6]

Another robustness check along the lines of Croushore (2010) is to examine whether the results change for samples with different ending dates. The intuition of this examination is to show what a researcher standing at different points in time would have found when undertaking the same exercise as Rudebusch and Williams. To test this, we look at differing ending dates for the sample. Do the

[6]See Stock and Watson (2003) and Ng and Wright (2013) for more on the deterioration since the 1990s of the power of the term spread as a forecasting variable.

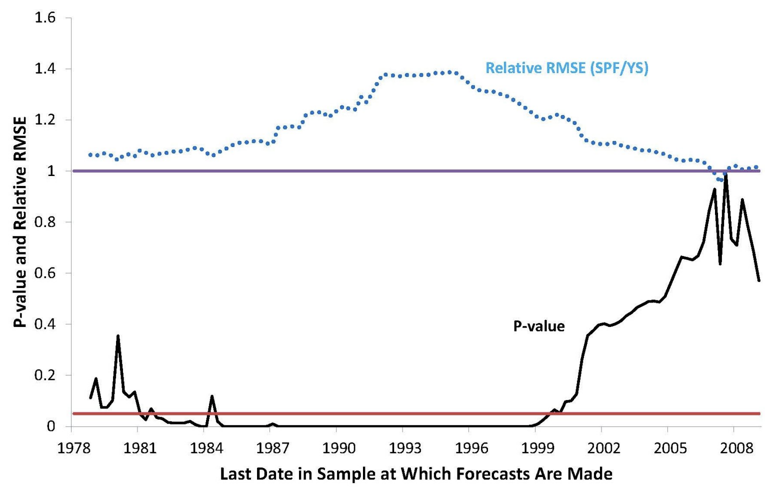
results change when we change the ending date? The results are consistent with the statistically significant superiority of the yield-spread probit over the SPF for four-quarter-ahead forecasts from samples that begin with the SPF survey of 1968:Q4 and end at almost any date from 1980:Q1 to 2012:Q1. Because the yield-spread probit forecast is uniformly superior to the SPF when we start the sample at 1968:Q4, and the differences are not statistically significant in just a few cases, we do not show the results here, but they are available from the authors upon request.

Because the effectiveness of using the yield spread to forecast a decline in real GDP seems to depend on the starting date of the sample, we next examine rolling windows for the forecast sample used in the evaluation. This method is often used in the forecasting literature; see Elliott and Timmermann (2008) for a discussion. Conceptually, the idea of examining forecasts in rolling windows is that we will more easily be able to observe changes in the structure of the economy—in this case, changes in the ability of the yield spread to forecast declines in real output.

Our results are shown in Figure 2, in which we again look at four-quarter-ahead forecasts, using 10-year windows, where the last SPF forecast date of the 10-year window is shown on the horizontal axis. Two lines are plotted: one showing the p-value of the significance of the difference between the two forecasts and the other showing the relative $RMSE$ of the forecasts. An $RMSE$ greater than 1 means that the yield-spread probit model has a lower $RMSE$ over the 10-year rolling window; an $RMSE$ of less than 1 means the SPF model has a lower $RMSE$. The figure shows that, for most 10-year windows, the yield-spread probit model is superior to the SPF and often significantly so. Toward the end of the sample, this becomes less true, but the SPF forecast is never significantly better than the yield-spread probit model and is better only in a few periods.

Rudebusch and Williams based their results on using yield-spread data beginning in 1955:Q1. We investigated the robustness of their results to that choice of starting dates and found no differences. Thus, the starting date of the yield-

Figure 2
P-Values and Relative RMSEs for 10-Year Rolling Windows



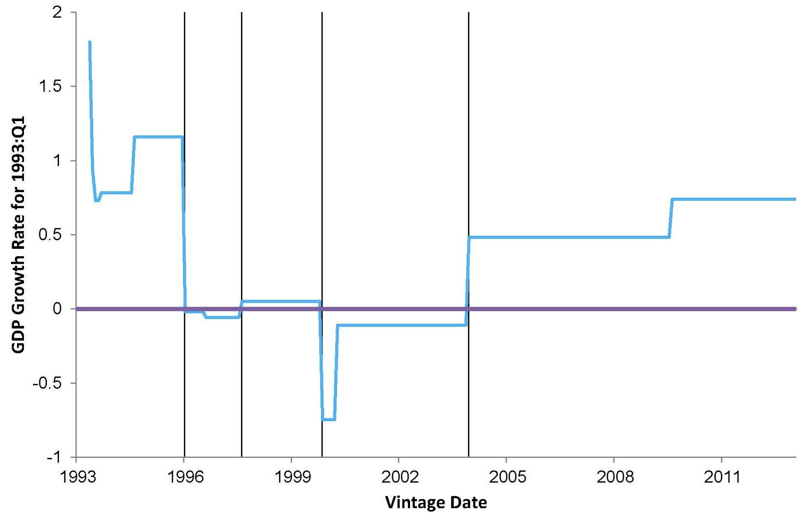
Note: The p-values in the bottom part of the figure show the significance of the difference between the yield-spread probit forecast and the SPF forecast based on $RMSE$. The top part of the figure shows the relative $RMSE$, defined as the $RMSE$ for the SPF forecast divided by the $RMSE$ for the yield-spread probit forecast. A relative $RMSE$ greater than 1 indicates that the yield-spread probit forecast has a lower $RMSE$.

spread data used in the probit model does not affect the results, which we are not showing here to conserve space.

C. Alternative Actuals

Another issue, which Rudebusch and Williams investigate a bit in their paper, is whether the choice of concept to use as "actual" matters. They use the first-final data, and all of our results so far in this paper have also used this concept. They also investigate the robustness of their results to choosing the initially released data and the final, revised data available to them when they wrote their paper. Because Croushore (2012) finds that results of forecast-bias tests can change substantially depending on the concept of "actual" that is used, we will explore the results in several dimensions: 1) using the latest available data that forecasters use in practice; and 2) using two additional alternative concepts of "actual" to see if that choice matters.

Figure 3
GDP Growth Rate for 1993:Q1

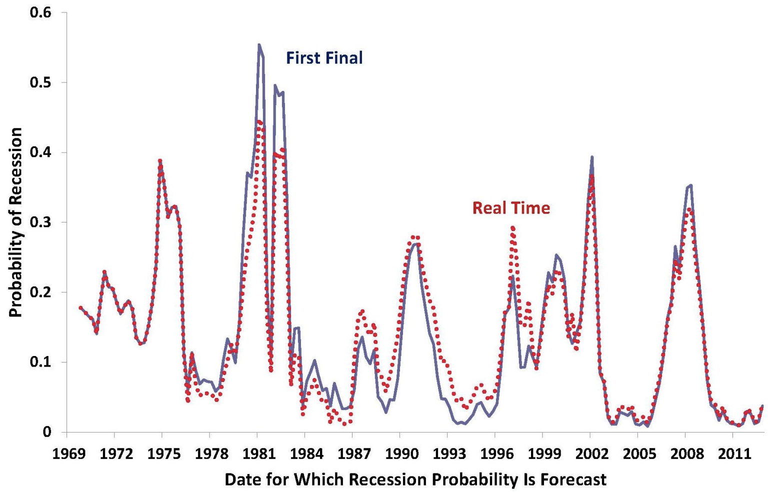


Note: The figure shows the growth rate of real GDP for 1993:Q1, as measured in the vintage shown on the horizontal axis. In 1993:Q1, there are four changes in the sign of the growth rate, from negative to positive or positive to negative, as the data are revised over time.

Rudebusch and Williams use the first-final data, taken as a data series, in calculating the $R1$ series to use in the yield-spread probit model and as actual $R1$ dates, but forecasters and researchers almost never use such data. Instead, common practice is to treat the latest available data as the best data available. So, someone in 1968:Q4 using the yield spread to forecast declines in real output over the next year is likely to estimate Equation (1) using the entire 1968:Q4 vintage of data on real output. The results based on latest available data may differ in several ways from results based on the first-final data used by Rudebusch and Williams. First, this approach might change which periods are defined as $R1$ recessions, as real GDP growth rates get revised and some switch from positive to negative, or vice versa. Second, the change in method to a period-by-period approach may lead to very different empirical results for the forecasting ability of the yield spread.

One result of this period-by-period real-time procedure is that some of the $R1$ recession dates change. For example, Figure 3 shows the growth rate of real GDP

Figure 4
Alternative R1 Recession Probabilities from
Four-Quarter-Ahead Yield-Spread Probit Model



Note: Each line shows the probability of a decline in real output four quarters in the future, using the yield-spread probit model. One line is based on first-final $R1$ dates, while the other line simulates a true real-time analysis, allowing the $R1$ dates to change when the data are revised.

for 1993:Q1 as someone observing the data at different dates (the vintage date shown on the horizontal axis) would view it. Vertical lines show when the real GDP growth rate changes sign. In the graph, you can see that the view of whether 1993:Q1 is an $R1$ recession quarter changes periodically over time. Other dates are also subject to changes in whether they are classified as $R1$ recession dates or not, depending on the vintage of data being used. Does this change the result of the yield-spread probit model? It could because the probit model is based on $R1$ dates as measured in real time, which are subject to change. In every benchmark NIPA revision from 1980 to 2009, real output growth changes from negative to positive, or vice versa, in at least one and as many as 10 quarters.

To investigate whether switches in the $R1$ recession dates matter, we look at the period-by-period real-time evolution of the recession forecasts using the vintage data at each date in the probit model, rather than just the sequence of first-final data vintages. Results are shown in Figure 4. The figure shows that the probability of a recession using the probit model based on the yield spread is slightly different if real-time data, and not first-final data, are used. The difference is that, in the real-time calculation, a forecaster using the yield spread would use the latest available real GDP data in running the model, whereas, in the first-final calculation, the forecaster ignores the latest available data but bases the model on the first-final data from the past. However, the differences between the two sets of forecasts are small enough that statistical tests are not likely to be strongly affected.

Suppose we rerun the exercise shown in Table 2, but this time using the real-time $R1$ recession dates rather than the first-final recession dates. We evaluate the $R1$ recession forecasts using the final vintage of real GDP data to determine the MAE, RMSE, and LPS.[7] The results are shown in Table 3. Comparing Table 2 to Table 3 shows only very small differences. So, the main Rudebusch-Williams results are robust to the use of real-time data instead of first-final data.

V. Summary and Conclusions

In this paper, we replicated the results of Rudebusch and Williams (2009) and found them to be robust in numerous dimensions. Our tests included extending the sample to include the 2007–09 recession, changing the starting date of the sample, changing the ending date of the sample, using rolling windows of data instead of an expanding sample, and using alternative measures of the "actual" value of real output. Our main finding is that using the yield-spread probit model to predict when real output will decline in a quarter produces forecasts that are superior those of the Survey of Professional Forecasters for horizons of three or four quarters.

There is some evidence, shown in Figure 2, that recently the SPF forecasters have been using the information in the yield spread in their recession forecasts. In the figure, we see that, in 10-year rolling windows since about 2005, the relative

[7]Note that using the first-final $R1$ recession dates in calculating MAE, RMSE, and LPS does not affect the table much at all; results are available from the authors upon request.

Table 3
Evaluation of Real-time Probability Forecasts:
May 2013 Data, Sample Ending in 2012

Probability Forecast	Full Sample MAE	Full Sample RMSE	Full Sample LPS	Post-1987 Sample MAE	Post-1987 Sample RMSE	Post-1987 Sample LPS
			Current-quarter R1 recession prediction			
SPF reported	**0.176†**	**0.269†**	**0.256†**	**0.150†**	**0.226†**	**0.194†**
Yield spread	0.224	0.342	0.393	0.199	0.313	0.353
			One-quarter-ahead R1 recession prediction			
SPF reported	0.219	**0.298**	**0.305**	**0.177**	**0.233**	**0.222**
Yield spread	**0.199**	0.319	0.362	0.187	0.308	0.349
			Two-quarter-ahead R1 recession prediction			
SPF reported	0.266	0.357	0.411	0.223	0.299	0.316
Yield spread	**0.209†**	**0.325**	**0.379**	**0.188†**	**0.297**	**0.312**
			Three-quarter-ahead R1 recession prediction			
SPF reported	0.274	0.366	0.446	0.236	0.319	0.362
Yield spread	**0.209†**	**0.322†**	**0.340†**	**0.180†**	**0.290**	**0.289†**
			Four-quarter-ahead R1 recession prediction			
SPF reported	0.279	0.370	0.455	0.239	0.325	0.375
Yield spread	**0.216†**	**0.327†**	**0.345†**	**0.182†**	**0.288†**	**0.278†**

Note: As in Table 1, bold numbers show the lowest value for each pair, comparing SPF reported value versus yield-spread probit value. Daggers indicate statistically significant differences at the 5 percent level (Diebold-Mariano test with HAC standard errors). The difference between these results and those shown in Table 2 is that the yield-spread probit model used here is based on real-time $R1$ dates rather than first-final dates.

RMSE has been close to 1, and the yield-spread probit model performs about as well as the SPF does. In 10-year windows since about 2000, the yield-spread probit model is no longer statistically significantly better than the SPF. But never is the SPF statistically significantly better than the probit model using the yield spread.

One unanswered question is whether the forecasts of a decline in real output made by the SPF participants are useful. One could argue that most professional forecasters focus their attention on point forecasts of a variety of macroeconomic variables and devote much less time to the probability forecasts in the SPF. Still, these forecasters are among the best in the country, as Croushore (1993) suggests, and their point forecasts are superior to even the most sophisticated econometric models, as Ang, Bekaert, and Wei (2007) show. So, if anyone can forecast recessions, it should be the forecasters in the SPF. But the fact that the SPF forecasts are inferior to a simple probit model using the yield spread is surprising and might lead people to have less faith in their quality.

REFERENCES

Ang, Andrew, Geert Bekaert, and Min Wei. 2007. "Do Macro Variables, Asset Markets, or Surveys Forecast Inflation Better?" *Journal of Monetary Economics*, 54: 1163–1212.

Bernard, Henri, and Stefan Gerlach. 1996. "Does the Term Structure Predict Recessions? The International Evidence." *BIS Working Paper Series 37*.

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

Croushore, Dean. 2010. "An Evaluation of Inflation Forecasts from Surveys using Real-Time Data." *B.E. Journal of Macroeconomics: Contributions*, 10(1).

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

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

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

Diebold, Francis X., and Glenn D. Rudebusch. 1991. "Turning Point Prediction with the Composite Leading Index: An Ex Ante Analysis." In *Leading Economic Indicators: New Approaches and Forecasting Records*, K. Lahiri and G.H. Moore, eds., 231–256. Cambridge University Press.

Elliott, Graham, and Allan Timmermann. 2008. "Economic Forecasting." *Journal of Economic Literature*, 46(1): 3–56.

Estrella, Arturo, and Gikas A. Hardouvelis. 1991. "The Term Structure as a Predictor of Real Economic Activity." *Journal of Finance*, 46(2): 555–576.

Lahiri, Kajal, George Monokroussos, and Yongchen Zhao. 2013. "The Yield Spread Puzzle and the Information Content of SPF Forecasts." *Economics Letters*, 118(1): 219–221.

Ng, Serena, and Jonathan H. Wright. 2013. "Facts and Challenges from the Great Recession for Forecasting and Macroeconomic Modeling." *Journal of Economic Literature*, 51: 1120–1154.

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. 1989. "New Indexes of Coincident and Leading Economic Indicators." In *NBER Macroeconomics Annual 1989*, Vol. 4, Olivier Jean Blanchard and Stanley Fischer, eds., 351–409. MIT Press.

Stock, James H., and Mark W. Watson. 1993. "A Procedure for Predicting Recessions with Leading Indicators: Econometric Issues and Recent Experience." In *Business Cycles, Indicators and Forecasting*, James H. Stock and Mark W. Watson, eds., 95–156. University of Chicago Press.

Stock, James H., and Mark W. Watson. 2003. "Forecasting Output and Inflation: The Role of Asset Prices." *Journal of Economic Literature*, 41: 788–829.

Zarnowitz, Victor. 1985. "Rational Expectations and Macroeconomic Forecasts." *Journal of Business and Economic Statistics*, 3: 293–311.

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*, James H. Stock and Mark W. Watson, eds., 11–94. University of Chicago Press.