



Realistic Evaluation of Real-Time Forecasts in the Survey of Professional Forecasters*

Tom Stark
Federal Reserve Bank of Philadelphia

May 28, 2010

Introduction

Each quarter, the Federal Reserve Bank of Philadelphia publishes the Survey of Professional Forecasters (SPF). This quarterly survey polls professional economists on their views about the economy over the next few years. Previously conducted by the National Bureau of Economic Research and the American Statistical Association, the survey provides forecasts for nominal GDP and the GDP price index, corporate profits, real GDP and its components, and a number of monthly business indicators, such as interest rates, housing starts, industrial production, and the consumer price index. The Philadelphia Fed has conducted this survey since the second quarter of 1990.

How accurate are the SPF forecasts? On first look, this seems like a straightforward question. The Philadelphia Fed posts the entire history of the survey's projections on its website, and it is easy to compare these projections with the historical data. On second look, as Dean Croushore and I have noted in our work over the last 10 years, macroeconomic historical data are revised often — and, in many cases, significantly.¹ Such revisions complicate the evaluation of forecasts.

In this note, I report on the survey's accuracy using alternative values of the historical realizations from the Philadelphia Fed's real-time data set for macroeconomists. I quantify the extent to which the data revisions matter and study the survey's performance relative to that of simple benchmark time-series models, estimated on the same real-time realizations that the panelists would have used at the time.

* The views expressed here are those of the author and do not necessarily reflect the views of the Federal Reserve Bank of Philadelphia or the Federal Reserve System. Tom Stark is the assistant director and manager of the Real-Time Data Research Center in the Philadelphia Fed's Research Department and may be contacted at tom.stark@phil.frb.org.

¹ See, for example, Croushore and Stark (2001) and Stark and Croushore (2002).

The major findings are:

- The survey's accuracy falls sharply at quarterly horizons beyond the first.
- Data revisions can have a large effect on the survey's accuracy; however, these revisions tend not to affect the survey's accuracy relative to that of the benchmark projections.
- The survey's projections easily outperform no-change forecasts for all variables except long-term interest rates.
- The survey's projections generally outperform the benchmark projections of univariate autoregressive time-series models at short horizons. At long horizons, relative accuracy depends on the variable forecast and the benchmark model considered.

I begin with a review of the timing of the survey's data set and the timing of the Philadelphia Fed's real-time data set. I then discuss my real-time forecast evaluation techniques and present the results.

The Timing of the SPF: What the Forecasters Know (and What They Do Not)

Real-time forecast evaluation requires paying careful attention to the data that the forecasters knew when they generated their projections. Such information is crucial in estimating and forecasting realistic comparison benchmark models. The Philadelphia Fed has maintained a steady schedule in conducting the SPF since the third-quarter survey of 1990.² We align the survey's schedule to the Bureau of Economic Analysis's (BEA) advance release of the data from the national income and product accounts. The BEA's report, issued in the latter part of January, April, July, and October, includes the first estimate of the historical realization of variables from the national accounts. The panelists return their projections before the BEA releases, one month later, the first revision to the advance estimate.³ Thus, the panelists' information set includes the advance estimate of the data from the national accounts. We send our questionnaires to the forecasters after these data are released to the public (Figure 1, bottom panel).⁴

The survey's timing is somewhat less certain prior to the second quarter of 1990. However, a recent internal analysis by Calvin Price, an analyst in the Philadelphia Fed's Real-Time Data Research Center, suggests that the NBER/ASA's schedule was about the same as the current SPF schedule. Price

² Our first survey was that of the second quarter of 1990. However, this survey was conducted after the fact because we had not yet received all the information from the NBER/ASA that we needed to conduct it in real time.

³ See Nakamura and Stark (2007) for a detailed discussion of the timing of revisions to the data from the national income and product accounts.

⁴ The only exception is the survey's measure of after-tax corporate profits. The BEA releases this variable with at least a one-month lag relative to the survey's other variables from the national accounts. For additional information, see the documentation for this variable in the Philadelphia Fed's real-time data set, at http://www.philadelphiafed.org/research-and-data/real-time-center/real-time-data/data-files/NCPROFATW/specific_documentation_NCPROFATW.pdf.

compared the latest-available historical observation in the survey's data set with the values as they appear in the Philadelphia Fed's real-time data set and found a close correspondence, particularly since 1985, the period that I study here.⁵

Some uncertainties about timing exist in the survey's projections for monthly variables not reported in the national accounts. It is very reasonable to assume that the survey's participants know the values of interest rates and labor-market variables for the first month of the first quarter of the quarterly forecast horizon.⁶ In contrast, the panelists generally do not know the latest monthly values of housing starts, industrial production, and the consumer price index when they form their projections (Figure 1, bottom panel).

An Information Set for the Benchmark Models: The Philadelphia Fed's Real-Time Data Set for Macroeconomists

Realistic real-time forecast evaluation means that I match the data used to estimate and forecast benchmark models with the data the survey's panelists would have used at the time they submitted their projections. The Philadelphia Fed's real-time data set contains vintages — or snapshots — of the historical observations as they appeared in the past. Nearly every variable in the survey has a counterpart variable in the real-time data set.⁷

For survey variables reported in the national income and product accounts, I estimate benchmark models on the vintages of data that were available to the public in the middle of the quarter. For the remaining variables, I use the monthly vintage for the first or second month of the quarter, depending on whether government statistical agencies release their data before or after the survey's mid-quarter deadline (Figure 1, top panel).

Methods and Benchmark Models

I measure forecast accuracy with the root-mean-square error, defined for the quarterly horizon τ and indexed by the measure of the realization (r) as

⁵ These results are available on request.

⁶ See Stark (2000) for a study of the benefits of using high-frequency information to condition one's forecast for quarterly observations.

⁷ One notable exception is CPI inflation. The real-time data set contains only a limited number of vintages for this variable. Moreover, the timing of the data releases and our vintages can differ by one month from the survey's timing.

$$rmse(\tau, r) = \sqrt{\frac{1}{T} \sum_{t=1}^T (\hat{\varepsilon}_{t+\tau|t}^{(r)})^2}, \quad \tau = 0, \dots, 4, \quad r = 1, \dots, 5,$$

where $\hat{\varepsilon}_{t+\tau|t}^{(r)} = y_{t+\tau}^{(r)} - \hat{y}_{t+\tau|t}$, $\hat{\varepsilon}_{t+\tau|t}^{(r)}$ is the τ -step-ahead forecast error, indexed by the measure of the realization (r), $y_{t+\tau}^{(r)}$ is the r -indexed realization, and $\hat{y}_{t+\tau|t}$ is the projection for period $t + \tau$ made in period t . I define five alternative measures of the realizations. The first realization is the value that government statistical agencies report in their initial releases of the data. The second is the value that the statistical agencies publish one quarter later. The third and fourth measures are those published five quarters and nine quarters later, respectively. The fifth measure comes from the data as they are reported at the time I made the computations. The realizations are subject to increasing degrees of revision as we move from the first measure to the fifth.

The forecasts of univariate time-series models provide benchmark comparisons. Let T and $V(T)$ define the date on which we conducted a survey and the vintage of data available at that date, respectively. The SPF panelists provide projections for the current quarter (T) and the following four quarters ($T + 1, \dots, T + 4$).

The first benchmark model is the no-change – or random walk – specification given by

$$\hat{y}_{T+\tau|T} = y_{T-1}^{(V(T))}, \quad \tau = 0, \dots, 4,$$

where $\hat{y}_{T+\tau|T}$ is the projection for a stationary transformation of a survey variable, and $y_{T-1}^{(V(T))}$ is the historical realization for quarter $T - 1$, as reported in the vintage of $V(T)$. This forecasting rule says that the benchmark projection for the current quarter ($\tau = 0$) and beyond ($\tau > 0$) is the same as the historical realization for the previous quarter, as recorded in the vintage of $V(T)$. For each quarterly survey of date T , I generate a new forecast for the horizons $T + \tau$, $\tau = 0, \dots, 4$. This procedure yields a sequence of real-time, no-change projections given by $\{\hat{y}_{T+\tau|T}\}_{\tau=0, T=\underline{T}}^{\tau=4, T=\bar{T}}$, where \underline{T} and \bar{T} are the first and last surveys of my sample period. I construct analogous sequences for each of the following benchmark models.

The second model is a one-period univariate autoregression given by

$$y_t^{(V(T))} = \phi_0 + \sum_{j=1}^{L(V(T))} \phi_j y_{t-j}^{(V(T))} + u_t, \quad t = T - 60, \dots, T - 1$$

where ϕ_j , $j = 1, \dots, L(V(T))$ are the parameters and u_t is the disturbance. For each survey, I re-estimate the parameters on a rolling fixed window of 60 observations, using observations from the vintage $V(T)$.⁸ The lag length, $L(V(T))$, chosen according to the Akaike information criterion, is also re-estimated each quarter. I compute the out-of-sample forecasts for this model indirectly, using the standard chain rule. This is the IAR model (indirect autoregression).

The third model modifies the IAR by using a different equation at each horizon. Each equation is evaluated directly to produce an out-of-sample forecast for that horizon. Research by Schorfheide (2005) and Marcellino, Stock and Watson (2006) suggests that a direct autoregression (DAR) may reduce the root-mean-square error under the presence of certain types of model misspecification. The DAR model at horizon $\tau = 0, \dots, 4$ is given by

$$y_t^{(V(T))} = \phi_0^{(\tau)} + \sum_{j=0}^{L(\tau, V(T))-1} \phi_{1+j}^{(\tau)} y_{t-1-\tau-j}^{(V(T))} + u_t^{(\tau)}, \quad t = T - 60, \dots, T - 1, \quad \tau = 0, \dots, 4$$

where $\phi_{1+j}^{(\tau)}$, $j = 0, \dots, L(\tau, V(T)) - 1$, $L(\tau, V(T))$, and $u_t^{(\tau)}$ are the horizon-indexed parameters, lag length, and disturbance, respectively. As above, I re-estimate the parameters and lag length quarterly on a rolling fixed window of 60 observations, using the real-time data of vintage $V(T)$.

The fourth benchmark model augments the DAR with monthly predictors (DARM). For survey variables whose historical values are reported at a monthly frequency, a direct autoregression augmented with monthly predictors adds, at most, three monthly predictors, according to

$$y_t^{(V(T))} = \phi_0^{(\tau)} + \sum_{j=0}^{L(\tau, V(T))-1} \phi_{1+j}^{(\tau)} y_{t-1-\tau-j}^{(V(T))} + \gamma_1 m1_{t-\tau}^{V(T)} + \gamma_2 m3_{t-\tau-1}^{V(T)} + \gamma_3 m2_{t-\tau-1}^{V(T)} + u_t^{(\tau)}, \quad t = T - 60, \dots, T - 1,$$

$$\tau = 0, \dots, 4,$$

⁸ Swanson (1998) and Pesaran and Timmermann (2007) discuss the benefits of fixed windows.

where $m1_{t-\tau}^{V(T)}$, $m3_{t-\tau-1}^{V(T)}$, and $m2_{t-\tau-1}^{V(T)}$ are the monthly values of the (quarterly-average) dependent variable for the first month of quarter $t - \tau$ and the third and second months of quarter $t - \tau - 1$, respectively, as they are reported in the vintage $V(T)$. The variable $m1_t^{V(T)}$ represents the monthly value of the dependent variable in the first month of the survey quarter. I include this variable only when its public release comes before the survey's mid-quarter deadline.

How Accurate Are the SPF's Projections?

Tables 1 – 4 report the survey's root-mean-square errors for quarter-over-quarter real output growth and inflation, the unemployment rate, and the rate on 10-year Treasury bonds.⁹ The sample period is 1985 Q1 to 2007 Q4. Each table shows the quarterly forecast horizon, the corresponding root-mean-square error of the SPF, and the ratio of the SPF's root-mean-square error to that of the benchmark models. Two-sided p-values for the Diebold-Mariano (1995) test for the equality of mean-square errors between the SPF and each benchmark appear in parentheses.¹⁰

Real Output Growth and Other Real Variables from the National Accounts. The results of Table 1 underscore the difficulty the forecasters face in predicting real GDP. I find large root-mean-square errors at all horizons beyond the first (column SPF).¹¹ The root-mean-square errors range from 1.40 percent to 1.92 percent at the first horizon, depending on the measure of the realizations. They rise to a range of 1.78 percent to 2.18 at the last horizon. Figure 2.A maps these results to confidence intervals around the survey's latest forecast. The shaded areas show intervals for alternative ranges covering 25, 50, 65, 75, and 80 percent confidence. The wide intervals reflect large root-mean-square errors and highlight the degree of uncertainty surrounding forecasts for economic growth.¹²

⁹ Results for nearly all variables in the survey (updated following each quarterly survey), additional measures of forecast accuracy, and additional sample periods are available at: www.philadelphiafed.org/research-and-data/real-time-center/survey-of-professional-forecasters/data-files/error-statistics.cfm.

¹⁰ The p-values use a Bartlett window with a truncation lag equal to the forecast step ($\tau = 0, \dots, 4$) plus four. In most cases, the Harvey-Leybourne-Newbold (1997) correction and alternative truncation lags yield qualitatively similar results.

¹¹ Real output is defined as fixed-weighted real GNP in the surveys conducted before 1992 Q1, fixed-weighted real GDP in the surveys from 1992 Q1 to 1995 Q4, and chain-weighted real GDP in the surveys thereafter. The corresponding benchmark forecasts and measures of realizations are defined similarly.

¹² I assume that the forecast errors at each horizon are distributed normally with a mean of zero and a variance given by the root-mean-square error. A $(1 - \alpha)$ percent confidence interval is constructed using the formula given by

$\hat{y}_{T+\tau|T} \pm Z_{\alpha/2} \text{rmse}(\tau)$, where $Z_{\alpha/2}$ is the $100(\alpha/2)$ th percentile point of the standard normal distribution.

Data revisions have a pronounced effect on the accuracy of the survey's real GDP projections: The survey's forecasts become more inaccurate as the data are revised over time. The RMSE for the current-quarter forecast ($\tau = 0$) rises from 1.40 percent for initial-release realizations to 1.68 percent for realizations measured just one quarter later. I find even larger RMSEs with realizations measured five quarters later (1.84 percent) and nine quarters later (1.87 percent). Figure 2.B summarizes the results over all horizons by plotting the RMSEs as a function of the realization. At all horizons, the RMSE rises as the data are subject to additional revision. This finding suggests that the forecasters are better at predicting the early releases of real GDP than they are at predicting the later releases.

Interestingly, this result stands out among the survey's remaining real variables from the national accounts. For these variables, I find that estimated RMSEs can be sensitive to the measure of the realization. The RMSEs do not, however, necessarily rise over time. For example, the estimated RMSE for quarter-over-quarter growth in real personal consumption expenditures generally falls over time (not shown, but available on the Philadelphia Fed's website), suggesting that estimated accuracy rises as the data on personal consumption expenditures are revised.

The survey's projections for real output growth easily outperform the no-change forecasts by 20 to 30 percentage points of root-mean-square error (Table 1, column SPF/NC). The relative RMSEs against the no-change projection are well below unity at all horizons. The ease with which the SPF outperforms the no-change projection – which also holds for the survey's other real variables from the national accounts – reflects the existence of low persistence in the quarterly growth rates of real variables in the U.S. data. Under such conditions, the no-change model for growth performs poorly.¹³

The survey's projections generally outperform the projections of the autoregressive models at the shortest horizon (columns SPF/IAR and SPF/DAR). Beyond that point, the relative RMSEs approach unity, indicating that it is hard to distinguish the SPF forecasts for real output from the benchmark projections. Interestingly, the relative performance of the SPF at longer horizons is quite good for some other real variables from the national accounts, such as real business fixed investment and real federal government consumption and gross investment (not reported here, but available on the Philadelphia Fed's website).

The relative performance of the SPF is not much affected by revisions to the data. In most instances, when the survey produces a root-mean-square error lower than that of a benchmark, the same holds under alternative measures of the historical data. However, the margin by which one forecast

¹³ The SPF's relative advantage over the no-change model does not extend to all variables in the survey. Examples (not shown here, but available on the Philadelphia Fed's website) include the projections for housing starts and long-term interest rates. For these variables, it can be hard to distinguish statistically between the survey's projections and those of the no-change benchmark.

outperforms another can change as the data are revised. For example, the SPF projections for output outperform the autoregressive benchmarks by nearly 20 percentage points of RMSE at the current-quarter horizon when initial-release values measure the realizations (Table 1, Panel 1). The margin shrinks considerably under subsequent measures of the realizations (Panels 2 to 5).

On the whole, across the survey's forecasts for real variables from the national accounts, I find that data revisions often affect absolute measures of accuracy. Data revisions tend not to change the survey's accuracy relative to the benchmarks. Faust and Wright (2007) report similar results on the relative accuracy of the forecasts for real output from the Federal Reserve Board's Greenbook.

Inflation. Table 2 reports the results for quarter-over-quarter inflation in the output-price index.¹⁴ The root-mean-square errors range from slightly below unity at short horizons to slightly above unity at the long horizons (column SPF). These estimates are smaller than the corresponding estimates for real output growth. Accordingly, the confidence intervals are somewhat tighter (Figure 3).

Revisions to the historical data have little effect on the root-mean-square error. At the current-quarter horizon ($\tau = 0$), they fall in the range of 0.83 percent to 0.86 percent over all measures of the realizations. This tight range contrasts sharply with the results for real output, which were quite sensitive to revisions.

The survey outperforms the benchmark models at short horizons, regardless of the measure of the realization (columns S/NC, S/IAR, and S/DAR). The root-mean-square error ratios are substantially below unity at these horizons. At long horizons, the survey continues to outperform the autoregressive benchmarks but loses much of its relative advantage over the no-change forecast.

Unemployment. Historically, the survey has done quite well in predicting unemployment. The root-mean-square errors range from 0.12 percentage point in the current-quarter to 0.58 percentage point at the four-quarter-ahead horizon (Table 3, column SPF).¹⁵ The corresponding confidence intervals around the survey's latest projection, plotted in Figure 4, are quite narrow at the shortest horizons, but they expand rapidly at the longer horizons.

¹⁴ This variable is defined as the quarter-over-quarter percent change (annualized percentage points) in the implicit price deflator for GNP in surveys conducted prior to 1992 Q1, the implicit deflator for GDP in the surveys from 1992 Q1 to 1995 Q4, and the chain-weighted price index in the surveys thereafter. The corresponding benchmark forecasts and realizations are defined similarly.

¹⁵ The unemployment rate is subject to very minor revisions. Such revisions have little effect on my computations. The table reports the results for the realizations taken from the latest vintage of data available at the time I made the calculations.

The SPF easily outperforms the benchmark no-change and autoregressive forecasts, nearly 25 to 40 percentage points of root-mean-square error (columns S/NC, S/IAR, and S/DAR). However, a sizable amount of the survey's relative accuracy reflects the panelists' use of within-quarter information on the monthly unemployment rate. The last column in Table 3 quantifies the effect by reporting how the survey compares with the DARM benchmark projection, which incorporates the latest information on monthly unemployment. The survey's relative advantage shrinks considerably (column SPF/DARM).

Long-Term Interest Rates. The SPF has a mixed record on predicting the path of long-term interest rates (Table 4). The root-mean-square error for the rate on 10-year Treasury bonds rises from just 0.16 percentage point at the current-quarter horizon to nearly 1.0 percent at the four-quarter-ahead horizon. This large decline in accuracy produces confidence intervals that widen considerably with the forecast horizon (Figure 5).

The survey generally outperforms the no-change and autoregressive DAR and IAR benchmark models at the shortest horizons. Much of the survey's relative advantage at these horizons reflects the panelists' use of high-frequency monthly observations in formulating their projections. Against the DARM, which incorporates the monthly observations, the SPF's relative advantage is much less than it is against the models that use only quarterly-average observations.

For predicting long-term interest rates at the longest horizons, the no-change benchmark performs as well as the SPF and better than the remaining benchmarks. Qualitatively similar results (not reported, but available on the Philadelphia Fed's website) characterize the survey's projections for the rate on Moody's AAA corporate bonds.

Closing Remarks

Evaluating projections from the Philadelphia Fed's Survey of Professional Forecasters is harder than it first appears. Revisions to the macroeconomic historical data give rise to questions about the appropriate values to use in computing the survey's forecast errors. Moreover, subtle issues concerning the survey's timing give rise to questions about appropriate benchmark comparisons to use in forecast evaluation exercises. In this note, I discussed the survey's timing, showed how to apply the Philadelphia Fed's real-time data set for realistic real-time forecast evaluation, and presented the results.

I find that the SPF forecasts quite well at short horizons and often outperforms the forecasts of benchmark univariate time-series models. However, forecast accuracy often deteriorates dramatically as the horizon lengthens. Forecast confidence intervals are, accordingly, extremely wide. This finding suggests that a large degree of uncertainty surrounds forecasts at long horizons.

The root-mean-square error statistics for a number of variables — in particular, real output and some of its components — are quite sensitive to revisions in the data. However, the survey's performance relative to the benchmark forecasts is not much affected by revisions. Results beyond those presented here and the data for the projections and measures of realizations are available on the Philadelphia Fed's website at: www.philadelphiafed.org/research-and-data/real-time-center/survey-of-professional-forecasters/data-files/error-statistics.cfm.

References

- Croushore, Dean, and Tom Stark. "A Funny Thing Happened on the Way to the Data Bank: A Real-Time Data Set for Macroeconomists," Federal Reserve Bank of Philadelphia *Business Review* (September/October 2000).
- Croushore, Dean, and Tom Stark. "A Real-Time Data Set for Macroeconomists," *Journal of Econometrics* 105 (2001), pp. 111-30.
- Diebold, Francis X. and Roberto S. Mariano. "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics* 13 (1995), pp. 253-63.
- Faust, Jon, and Jonathan H. Wright. "Comparing Greenbook and Reduced Form Forecasts Using a Large Realtime Dataset," National Bureau of Economic Research Working Paper: 13397 (2007).
- Harvey, David, Stephen Leybourne, and Paul Newbold. "Testing the Equality of Prediction Mean Square Errors," *International Journal of Forecasting* 13 (1997), pp. 281-91.
- Marcellino, Massimiliano, James H. Stock, and Mark W. Watson. "A Comparison of Direct and Iterated Multistep AR Methods for Forecasting Macroeconomic Time Series," *Journal of Econometrics* 135 (November/December 2006), pp. 499-526.
- Nakamura, Leonard I., and Tom Stark. "Mismeasured Personal Saving and the Permanent Income Hypothesis," Federal Reserve Bank of Philadelphia Working Paper 07-8 (February 2007).
- Pesaran, M. Hashem, and Allan Timmermann. "Selection of Estimation Window in the Presence of Breaks," *Journal of Econometrics*, 2007, pp. 134-61.
- Schorfheide, Frank. "VAR Forecasting Under Misspecification," *Journal of Econometrics* (September 2005), pp. 99-136.
- Stark, Tom. "Does Current-Quarter Information Improve Quarterly Forecasts for the U.S. Economy?," Federal Reserve Bank of Philadelphia Working Paper 00-2 (January 2000).
- Stark, Tom, and Dean Croushore. "Forecasting With a Real-Time Data Set for Macroeconomists," *Journal of Macroeconomics* 24 (2002), pp. 507-31.
- Swanson, Norman R. "Money and Output Viewed Through a Rolling Window," *Journal of Monetary Economics*, 1998, pp. 455-73.

**Table 1. SPF Root-Mean-Square Error Statistics for Output Growth,
1985 Q1 – 2007 Q4**

Horizon (τ)	SPF	SPF/NC	SPF/IAR	SPF/DAR	SPF/DARM
<i>1. Realizations Are the Values on Their Initial Release</i>					
0	1.40	0.681 (0.000)	0.806 (0.023)	0.806 (0.023)	NA
1	1.65	0.770 (0.001)	0.925 (0.446)	0.933 (0.485)	NA
2	1.76	0.782 (0.017)	0.986 (0.865)	0.985 (0.863)	NA
3	1.81	0.770 (0.024)	0.994 (0.912)	0.993 (0.890)	NA
4	1.78	0.736 (0.003)	0.981 (0.712)	0.948 (0.286)	NA
<i>2. Realizations Are the Values Available One Quarter After Their Initial Release</i>					
0	1.68	0.778 (0.000)	0.878 (0.076)	0.878 (0.076)	NA
1	1.83	0.826 (0.006)	0.925 (0.375)	0.936 (0.424)	NA
2	1.90	0.785 (0.014)	0.956 (0.575)	0.958 (0.589)	NA
3	1.94	0.758 (0.020)	0.969 (0.588)	0.963 (0.500)	NA
4	1.95	0.742 (0.002)	0.975 (0.613)	0.953 (0.299)	NA
<i>3. Realizations Are the Values Available Five Quarters After Their Initial Release</i>					
0	1.84	0.764 (0.000)	0.860 (0.082)	0.860 (0.082)	NA
1	2.00	0.810 (0.001)	0.921 (0.304)	0.928 (0.330)	NA
2	2.09	0.836 (0.032)	0.965 (0.597)	0.970 (0.656)	NA
3	2.09	0.755 (0.004)	0.958 (0.340)	0.951 (0.252)	NA
4	2.08	0.758 (0.003)	0.960 (0.366)	0.922 (0.099)	NA
<i>4. Realizations Are the Values Available Nine Quarters After Their Initial Release</i>					
0	1.87	0.805 (0.007)	0.878 (0.161)	0.878 (0.161)	NA
1	2.03	0.814 (0.003)	0.923 (0.309)	0.927 (0.318)	NA
2	2.14	0.839 (0.007)	0.975 (0.672)	0.977 (0.698)	NA
3	2.15	0.792 (0.016)	0.971 (0.500)	0.962 (0.334)	NA
4	2.18	0.784 (0.005)	0.989 (0.764)	0.943 (0.221)	NA
<i>5. Realizations Are the Values From the Vintage of 2010 Q2</i>					
0	1.92	0.789 (0.000)	0.911 (0.168)	0.911 (0.168)	NA
1	2.05	0.825 (0.019)	0.961 (0.547)	0.962 (0.548)	NA
2	2.15	0.857 (0.096)	1.008 (0.888)	1.010 (0.859)	NA
3	2.15	0.800 (0.017)	1.002 (0.969)	0.991 (0.809)	NA
4	2.18	0.824 (0.044)	1.020 (0.654)	0.963 (0.505)	NA

Notes. The table reports root-mean-square error statistics for projections for quarter-over-quarter real GNP/GDP growth (annualized percentage points) from the Survey of Professional Forecasters (column SPF) and the ratio of the SPF RMSE to that of the no-change benchmark forecast (SPF/NC), the IAR benchmark forecast (SPF/IAR), and the DAR benchmark forecast (SPF/DAR). Ratios below unity indicate that the SPF outperforms the benchmark model. The quarterly forecast horizon corresponds to the SPF forecast for the current quarter ($\tau = 0$) through the forecast that is four quarters ahead ($\tau = 4$). Each panel (1 – 5) shows the results for a different measure of the realization. The sample endpoint is restricted to facilitate comparisons across alternative measures of the realizations. The numbers in parentheses are two-sided p-values for the Diebold-Mariano (1995) test of mean-square-error equality. The survey uses real GNP before 1992 Q1 and real GDP thereafter. The realizations and benchmark projections are defined similarly.

Table 2. SPF Root-Mean-Square Error Statistics for Output-Price Inflation, 1985 Q1 – 2007 Q4

Horizon (τ)	SPF	SPF/NC	SPF/IAR	SPF/DAR	SPF/DARM
<i>1. Realizations Are the Values on Their Initial Release</i>					
0	0.86	0.750 (0.002)	0.793 (0.006)	0.793 (0.006)	NA
1	0.95	0.723 (0.001)	0.813 (0.024)	0.815 (0.031)	NA
2	1.04	0.873 (0.125)	0.891 (0.180)	0.888 (0.126)	NA
3	1.15	1.011 (0.903)	0.888 (0.186)	0.842 (0.141)	NA
4	1.20	0.938 (0.427)	0.826 (0.062)	0.740 (0.065)	NA
<i>2. Realizations Are the Values Available One Quarter After Their Initial Release</i>					
0	0.86	0.722 (0.000)	0.778 (0.001)	0.778 (0.001)	NA
1	0.95	0.712 (0.000)	0.803 (0.009)	0.799 (0.013)	NA
2	1.02	0.816 (0.028)	0.868 (0.081)	0.863 (0.045)	NA
3	1.14	0.979 (0.803)	0.898 (0.212)	0.841 (0.157)	NA
4	1.20	0.962 (0.666)	0.846 (0.079)	0.760 (0.096)	NA
<i>3. Realizations Are the Values Available Five Quarters After Their Initial Release</i>					
0	0.86	0.718 (0.000)	0.798 (0.000)	0.798 (0.000)	NA
1	0.95	0.728 (0.000)	0.824 (0.001)	0.820 (0.003)	NA
2	1.02	0.860 (0.100)	0.880 (0.055)	0.886 (0.071)	NA
3	1.13	0.941 (0.358)	0.858 (0.064)	0.820 (0.122)	NA
4	1.15	0.907 (0.357)	0.800 (0.016)	0.717 (0.045)	NA
<i>4. Realizations Are the Values Available Nine Quarters After Their Initial Release</i>					
0	0.84	0.718 (0.000)	0.820 (0.001)	0.820 (0.001)	NA
1	0.91	0.730 (0.000)	0.858 (0.003)	0.859 (0.009)	NA
2	0.97	0.900 (0.274)	0.930 (0.261)	0.934 (0.351)	NA
3	1.10	0.973 (0.723)	0.884 (0.136)	0.842 (0.219)	NA
4	1.08	0.867 (0.231)	0.791 (0.029)	0.706 (0.057)	NA
<i>5. Realizations Are the Values From the Vintage of 2010 Q2</i>					
0	0.83	0.765 (0.004)	0.809 (0.025)	0.809 (0.025)	NA
1	0.94	0.790 (0.002)	0.833 (0.028)	0.841 (0.072)	NA
2	1.03	0.901 (0.261)	0.871 (0.156)	0.879 (0.162)	NA
3	1.16	0.987 (0.905)	0.861 (0.089)	0.807 (0.105)	NA
4	1.16	0.942 (0.606)	0.794 (0.048)	0.702 (0.068)	NA

Notes. The table reports SPF root-mean-square error statistics and ratios for projections for quarter-over-quarter growth in the GNP/GDP price index (annualized percentage points). See Table 1 for additional notes. The price index is that for GNP before 1992 Q1 and GDP thereafter. The realizations and benchmark projections are defined similarly.

Table 3. SPF Root-Mean-Square Error Statistics for the Unemployment Rate, 1985 Q1 – 2007 Q4

Horizon (τ)	SPF	SPF/NC	SPF/IAR	SPF/DAR	SPF/DARM
<i>Realizations Are the Values From the Vintage of 2010 Q2</i>					
0	0.12	0.576 (0.002)	0.600 (0.000)	0.600 (0.000)	0.853 (0.043)
1	0.25	0.693 (0.016)	0.703 (0.000)	0.722 (0.000)	0.823 (0.035)
2	0.36	0.723 (0.017)	0.691 (0.000)	0.710 (0.000)	0.803 (0.007)
3	0.47	0.764 (0.009)	0.712 (0.001)	0.743 (0.007)	0.810 (0.017)
4	0.58	0.774 (0.004)	0.713 (0.003)	0.741(0.017)	0.791 (0.033)

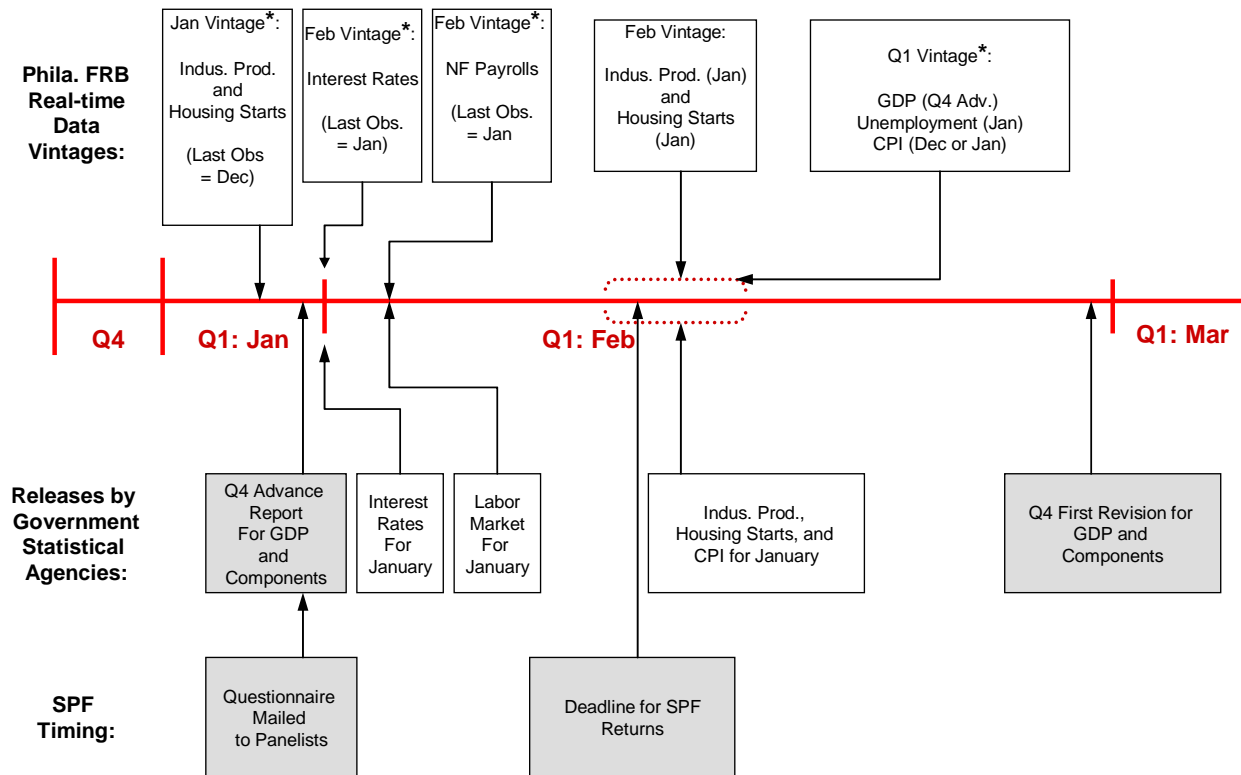
Notes. The table reports root-mean-square error statistics for projections for the quarterly average unemployment rate (percentage points) from the Survey of Professional Forecasters (column SPF) and the ratio of the SPF RMSE to that of the no-change benchmark forecast (SPF/NC), the IAR benchmark forecast (SPF/IAR), the DAR benchmark forecast (SPF/DAR), and the DARM benchmark forecast (SPF/DARM). The benchmark DARM model augments the DAR specification with monthly observations for the first month of the current quarter and the two most recent lagged values, corresponding to the third and second months of the previous quarter. Ratios below unity indicate that the SPF outperforms the benchmark model. The quarterly forecast horizon corresponds to the SPF forecast for the current quarter ($\tau = 0$) through the forecast that is four quarters ahead ($\tau = 4$). Because revisions to quarterly-average unemployment are small, all statistics use the latest available data as they appear in the vintage of 2010 Q2. The numbers in parentheses are two-sided p-values for the Diebold-Mariano (1995) test of mean-square-error equality.

Table 4. SPF Root-Mean-Square Error Statistics for the Rate on 10-Year Treasury Bonds, 1993 Q1 – 2007 Q4

Horizon (τ)	SPF	SPF/NC	SPF/IAR	SPF/DAR	SPF/DARM
<i>Realizations Are the Values From the Vintage of 2010 Q2</i>					
0	0.16	0.415 (0.000)	0.427 (0.000)	0.427 (0.000)	0.787 (0.009)
1	0.53	0.877 (0.048)	0.841 (0.006)	0.846 (0.013)	1.005 (0.920)
2	0.74	0.971 (0.615)	0.887 (0.155)	0.836 (0.062)	0.986 (0.815)
3	0.89	0.996 (0.936)	0.863 (0.104)	0.833 (0.071)	0.931 (0.392)
4	0.98	1.021 (0.733)	0.827 (0.063)	0.817 (0.084)	0.867 (0.144)

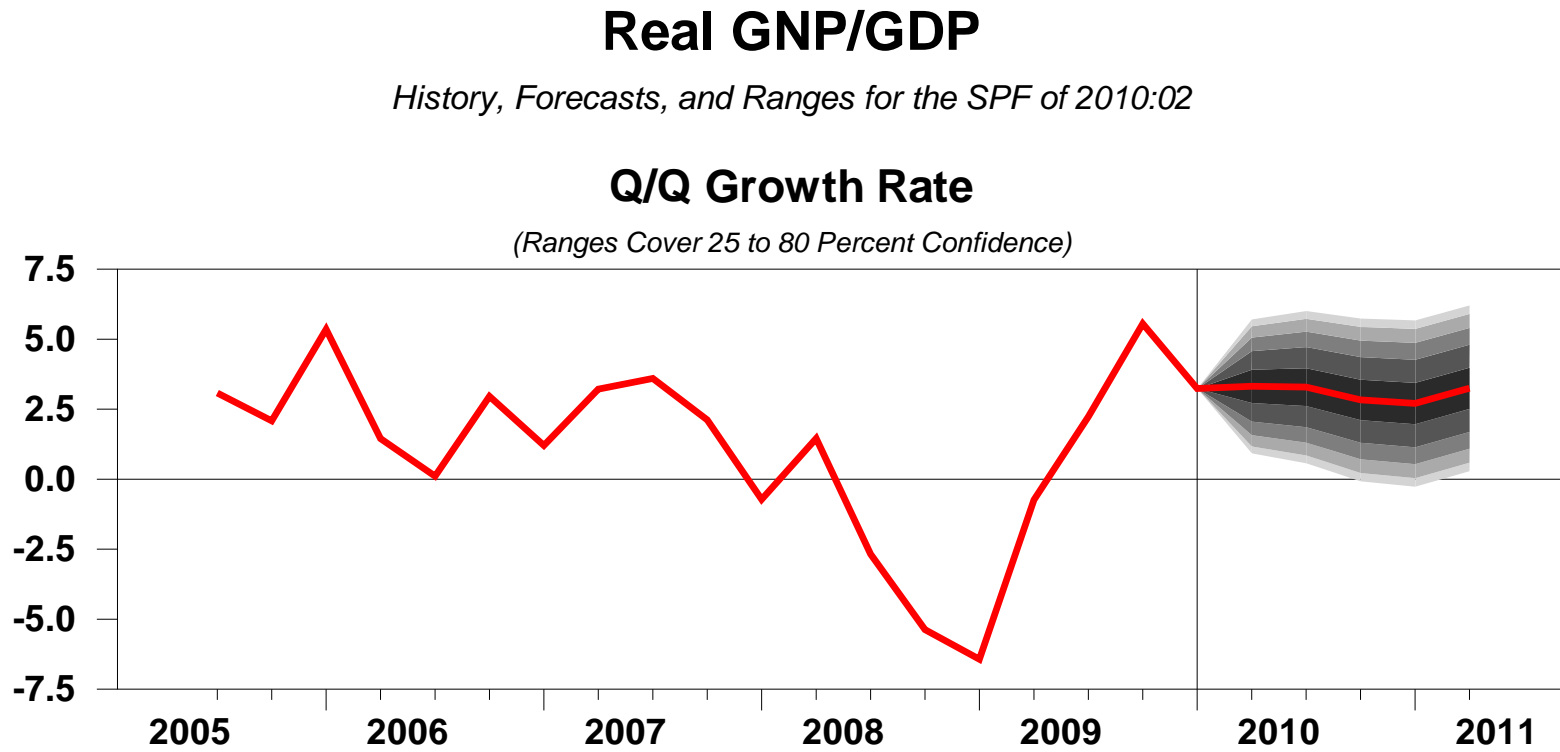
Notes. The table reports root-mean-square error statistics and ratios for the projections for the quarterly-average rate on 10-year Treasury bonds. See Table 3 for additional notes. The benchmark DARM model augments the DAR specification with the monthly observation for the first month of the current quarter. Ratios below unity indicate that the SPF outperforms the benchmark model. The numbers in parentheses are two-sided p-values for the Diebold-Mariano (1995) test of mean-square-error equality.

Figure 1. Relative Timing of the SPF, Statistical Releases, and Information Sets: First-Quarter Survey



Notes. The bottom panel shows the relative timing of major data releases, the date on which the questionnaire is released, and the deadline for returns. The top panel shows the dates on which we collect vintages of real-time data and the last observation included. Asterisks (*) denote vintages used. Industrial production, housing starts, and the CPI are released near the survey deadline. The Q1 vintage for the CPI may or may not include the observation for January. A large number of returns arrive on the day of the deadline.

Figure 2.A.

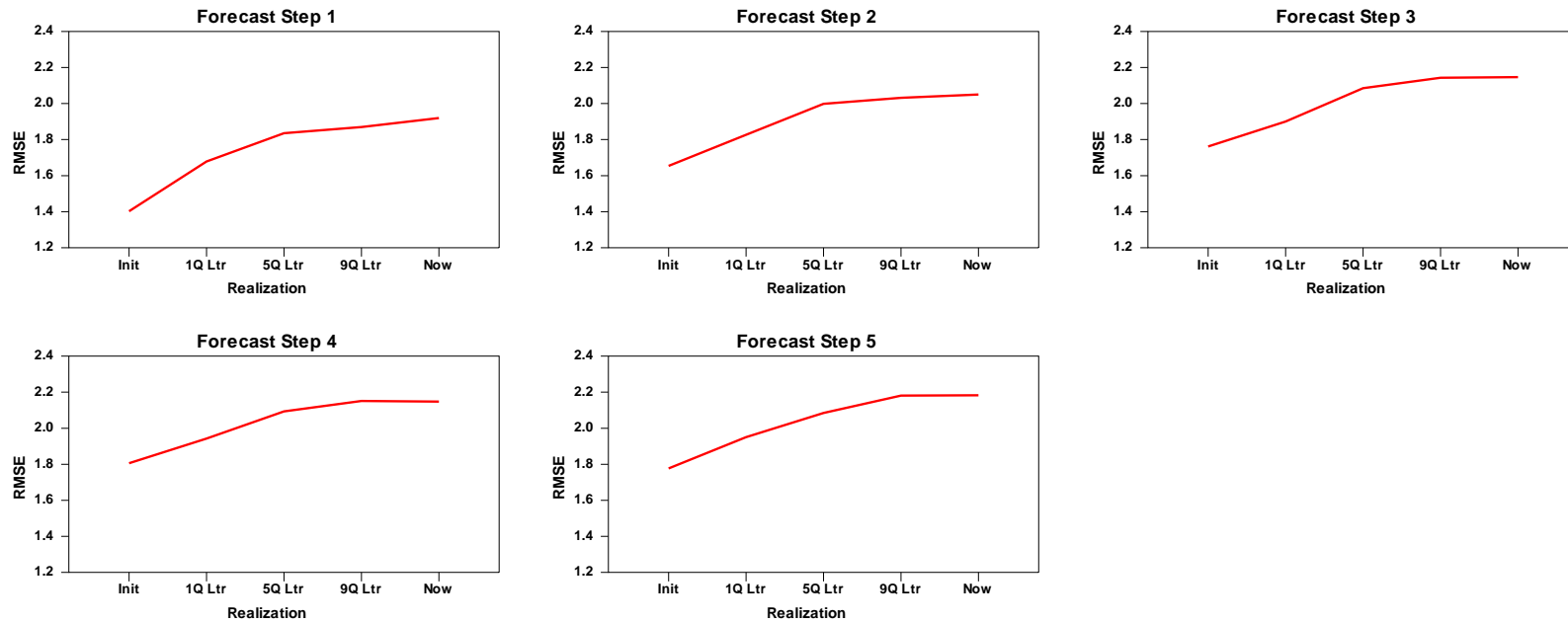


Ranges at each horizon use the $N(0, \text{MSE})$ density. The MSEs are based on the sample 85:01-08:04 and use the realization: Five Qtrs After Initial Release. Source: T.Stark, FRB Phila.

Figure 2.B.

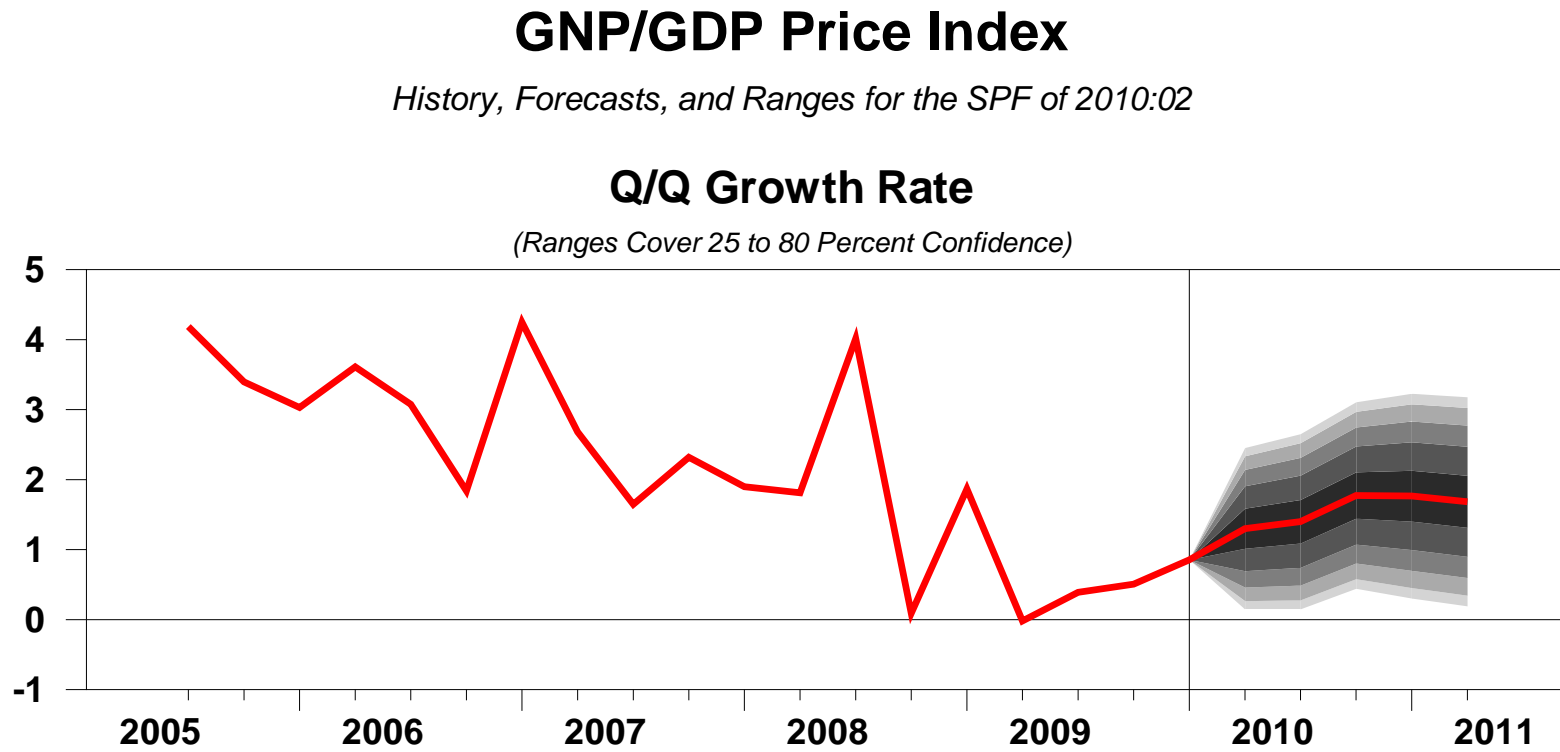
Root-Mean-Square Errors: 1985:01-2007:04

SPF Projections for Real GNP/GDP, Transformation: Q/Q Growth Rate



The RMSE is plotted against the realization used to compute it, from the value on initial release to the value as we now know it. Source: Tom Stark, FRB Philadelphia.

Figure 3.

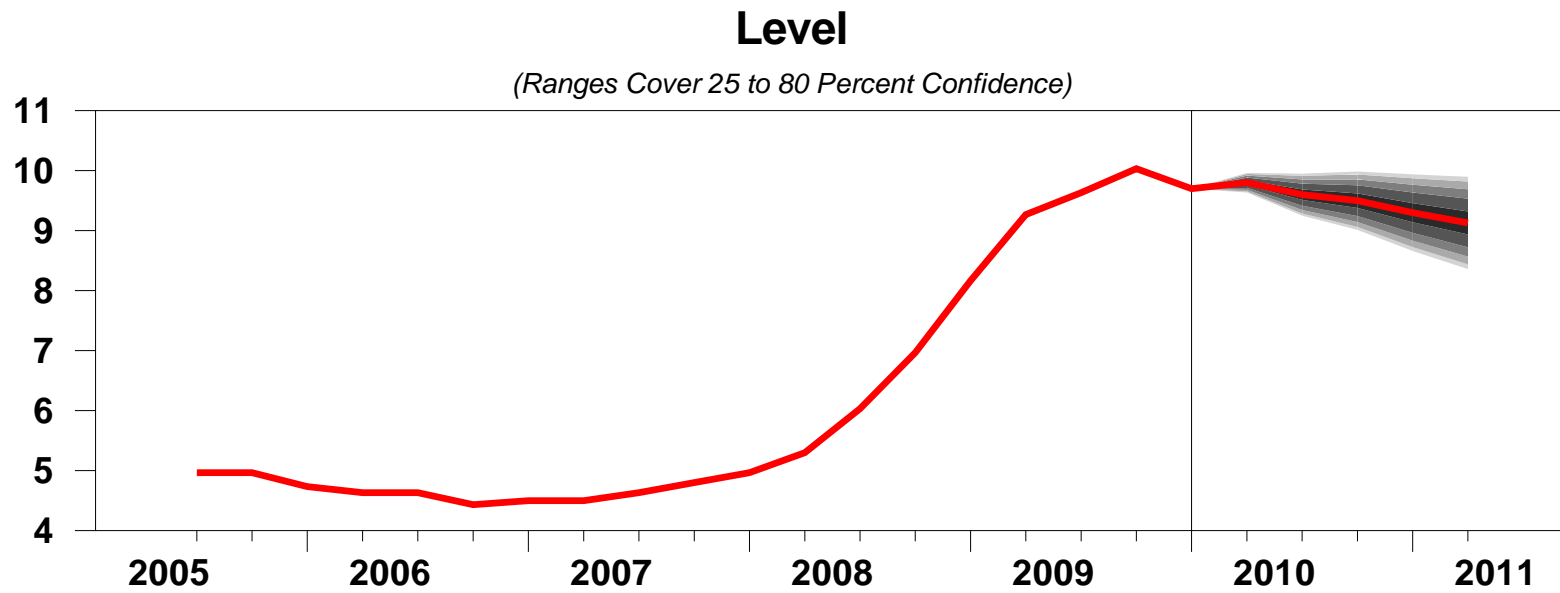


Ranges at each horizon use the $N(0, \text{MSE})$ density. The MSEs are based on the sample 85:01-08:04 and use the realization: Five Qtrs After Initial Release. Source: T.Stark, FRB Phila.

Figure 4.

Unemployment Rate [QA, PPs]

History, Forecasts, and Ranges for the SPF of 2010:02



Ranges at each horizon use the $N(0, \text{MSE})$ density. The MSEs are based on the sample 85:01-08:04 and use the realization: Five Qtrs After Initial Release. Source: T.Stark, FRB Phila.

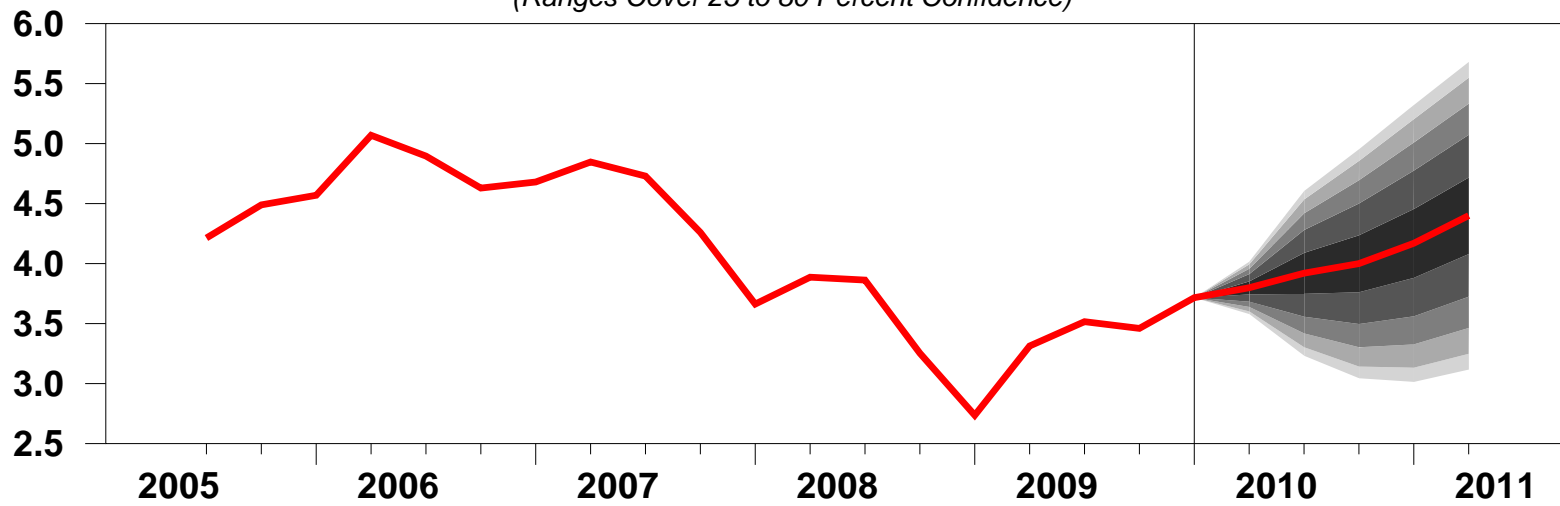
Figure 5.

10-Year T-Bond Rate [QA,PPs]

History, Forecasts, and Ranges for the SPF of 2010:02

Level

(Ranges Cover 25 to 80 Percent Confidence)



Ranges at each horizon use the $N(0, \text{MSE})$ density. The MSEs are based on the sample 93:01-08:04 and use the realization: Five Qtrs After Initial Release. Source: T.Stark, FRB Phila.