Error Statistics for the Survey of Professional Forecasters for CPI Inflation Rate

Release Date: 02/23/2021

Tom Stark
Assistant Vice President | Assistant Director
Real-Time Data Research Center
Economic Research Department
Federal Reserve Bank of Philadelphia

Source for Historical Realizations: Bureau of Labor Statistics via Haver Analytics

1. OVERVIEW.

This document reports error statistics for median projections from the Survey of Professional Forecasters (SPF), conducted since 1990 by the Federal Reserve Bank of Philadelphia. We provide the results in a series of tables and, in the PDF version of this document, a number of charts. The tables show the survey variable forecast and, importantly, the transformation of the data that we used to generate the statistics. (The transformation is usually a quarter-over-quarter growth rate, expressed in annualized percentage points. However, some variables, such as interest rates, the unemployment rate, and housing starts are untransformed and, thus, expressed in their natural units.)

The paragraphs below explain the format of the tables and charts and the methods used to compute the statistics. These paragraphs are general. The same discussion applies to all variables in the survey.

2. DESCRIPTION OF TABLES.

Tables 1A-1B report error statistics for various forecast horizons, sample periods, and choices of the real-time historical value that we used to assess accuracy. In each quarterly survey, we ask our panelists for their projections for the current quarter and the next four quarters. The current quarter is defined as the quarter in which we conducted the survey. Our tables provide error statistics separately for each quarter of this five-quarter horizon, beginning with the current quarter (denoted H = 1) and ending with the quarter that is four quarters in the future (H = 5). For each horizon, we report the mean forecast error [ME(S)], the mean absolute forecast error [MAE(S)], and the root-mean-square error [RMSE(S)]. All are standard measures of accuracy, though the academic literature generally places the most weight on the latter.

We define a forecast error as the difference between the historical value and the forecast. The mean error for each horizon is simply the average of the forecast errors at that horizon, constructed over the sample periods shown in Table 1A. Other things the same, a forecast with a mean error close to zero is better than one with a mean error far from zero. The mean absolute error is the sample average of the absolute value of the errors. Many analysts prefer this measure to the mean error because it does not allow large positive errors to offset large negative errors. In this sense, the mean absolute error gives a cleaner estimate of the size of the errors. Decision makers, however, may care not only about the average size of the errors but also about their variability, as measured by variance. Our last measure of accuracy is one that reflects the influence of the mean error and the variance of the error. The root-mean-square error for the SPF [RMSE(S)], the measure most often used by analysts and academicians, is the square root of the the average squared error. The lower the root-mean-square error, the more accurate the forecast.

2.1. Benchmark Models.

The forecast error statistics from the SPF are of interest in their own right. However, it is often more interesting to compare such statistics with those of alternative, or benchmark, forecasts. Tables 1A-1B report four such comparisons. They show the ratio of the root-mean-square error of the SPF forecast to that of four benchmark models. The benchmark models are statistical equations that we estimate on the data. We use the equations to generate projections for the same horizons included in the survey. In effect, we imagine standing back in time at each date when a survey was conducted and generating a separate forecast with each benchmark model. We do this in the same way that a survey panelist would have done using his own model.

Table 1A reports the root-mean-square-error ratios using as many observations as possible for each model. The number of observations can differ from model to model. We first compute the RMSE for each model. We then construct the ratio.

Table 1B reports RMSE ratios after we adjust the samples to include only the observations common to both models in the pair. Accordingly, the ratios reported in Table 1B may differ slightly from those of Table 1A, depending on the availability of sufficient real-time observations for estimating the benchmark models or for computing the errors of the SPF or benchmark forecasts. Table 1B also reports three two-sided p-values for each ratio. The p-values, corrected for the presence of heteroskedasticity and serial correlation in the time series of differences in squared forecast errors, are those for the test of equality of mean-square error between the SPF and the benchmark. The p-values are those for:

(1) The Diebold-Mariano statistic (July 1995, Journal of Business and Economic Statistics), using a uniform lag window with the truncation lag set to the forecast horizon minus unity. When the uniform lag window produces a nonpositive standard error, the Bartlett window is used.

(2) The Harvey-Leybourne-Newbold correction (1997, International Journal of Forecasting) to the Diebold-Mariano statistic.

(3) The Diebold-Mariano statistic, using a Bartlett lag window with the truncation lag increased four quarters beyond that of (1) and (2).

A RMSE ratio below unity indicates that the SPF consensus (median) forecast has a root-mean-square error lower than that of the benchmark. This means the SPF is more accurate. We now describe the benchmark models. The first is perhaps the simplest of all possible benchmarks: A no-change model. In this model, the forecast for quarter T, the one-step-ahead or current-quarter forecast, is simply the historical value for the prior quarter (T - 1). There is, in other words, no change in the forecast compared with the historical value. Moreover, the forecast for the remaining quarters of the horizon is the same as the forecast for the current quarter. We denote the relative RMSE ratio for this benchmark as RMSE(S/NC), using NC to indicate no change. The second and third benchmark models generate projections using one or more historical observations of the the variable forecast, weighted by coefficients estimated from the data. Such autoregressive (AR) models can be formulated in two ways. We can estimate one model to generate the forecasts at all horizons, using an iteration method to generate the projections beyond the current quarter (IAR), or we can directly estimate a new model for each forecast horizon (DAR). The latter formulation has been shown to reduce the bias in a forecast when the underlying model is characterized by certain types of misspecification. The root-mean-square error ratios are denoted RMSE(S/IAR) and RMSE(S/DAR), respectively.

The one- through five-step-ahead projections of the benchmark models use information on the quarterly average of the variable forecast. The latest historical observation is for the quarter that is one quarter before the quarter of the first projection in the horizon. In contrast, the panelists generate their projections with the help of additional information. They submit their projections near the middle of each quarter and hence have access to some monthly indicators for the first month of each quarter, when those data are released before the survey deadline. This puts the projections of panelists for some variables at an advantage relative to the corresponding benchmark projections. Moreover, the panelists may also examine the very recent historical values of such monthly indicators in forming their projections for quarterly averages. Such monthly statistical momentum represents an advantage not shared by the benchmark models, which use only quarterly averages. For survey variables whose observations are reported at a monthly frequency, such as interest rates, industrial production, housing starts, and unemployment, we estimate and forecast a fourth benchmark model, the DARM. This model adds recent monthly historical values to the specification of the DAR model. For the projections for unemployment, nonfarm payroll employment, and interest rates, we add the values of monthly observations, beginning with that for the first month of the first quarter of the forecast horizon. These values should be in the information set of the survey panelists at the time they formed their projections. In contrast, for variables such as housing starts and industrial production, we include only lagged values of monthly observations. For such variables, the panelists would not have known the monthly observation for the first month of the first quarter of the forecast horizon. In general, we find that adding monthly observations to the benchmark DAR models improves accuracy. Indeed, for the projections for interest rates and the unemployment rate, the accuracy of the benchmark DARM projections rivals that of the SPF projections.

2.2. Real-Time Data.

All benchmark models are estimated on a rolling, fixed window of 60 real-time quarterly observations. Lag lengths, based on either the Akaike information criterion (AIC) or the Schwarz information criterion (SIC), are re-estimated each period. The tables below indicate whether the lag length was was chosen by the AIC or SIC.

We would like to make the comparison between the SPF forecast and the forecasts of each benchmark as fair as possible. Therefore, we must subject the benchmark models to the same data environment the survey panelists faced when they made their projections. This is important because macroeconomic data are revised often, and we do not want the benchmark models to use a data set that differs from the one our panelists would have used. We estimate and forecast the benchmark models with real-time data from the Philadelphia Fed real-time data set, using the vintage of data that the survey panelists would have had at the time they generated their own projections. (For more information on the Philadelphia Fed real-time data set, go to www.philadelphiafed.org/econ/forecast/real-time-data/.)

An open question in the literature on forecasting is: What version or vintage of the data should we use to compute the errors? A closely related question is: What version of the data are professional forecasters trying to predict? Our computations take no strong position on these questions. In Tables 1A - 1B, we evaluate the projections (SPF and benchmark) with five alternative measures of the historical values, all from the Philadelphia Fed real-time data set. These measures range from the initial-release values to the values as we know them today. All together, we compute the forecast error statistics using the following five alternative measures of historical values:

(1) The initial or first-release value;
(2) The revised value as it appears one quarter after the initial release;
(3) The revised value as it appears five quarters after the initial release;
(4) The revised value as it appears nine quarters after the initial release;
(5) The revised value as it appears today.

Each measure of the historical value has advantages and disadvantages. The initial-release value is the first measure released by government statistical agencies. A forecaster might be very interested in this measure because it enables him to evaluate his latest forecast soon after he generated it. However, early releases of the data are often subject to large measurement error. Subsequent releases [(2) - (5)] are more accurate, but they are available much later than the initial release. As we go from the first measure to the fifth, we get more reliability, at the cost of higher delays in availability.

The last two columns in Table 1A report the number of observations that we used to compute the error statistics. Some observations are omitted because the data are missing in the real-time data set, such as occurred when federal government statistical agencies closed in late 1995.

2.3. Recent Projections and Realizations.

Tables 2 to 7 provide information on recent projections and realizations. They show how we align the data prior to computing the forecast errors that form the backbone of the computations in Tables 1A - 1B. Any error can be written as the equation given by error = realization - forecast. For our computations, we must be more precise because, for each projection (SPF and benchmarks), we have different periods forecast (T) different forecast horizons (h), and several measures of the realization (m). Thus, we can define the forecast error more precisely as

$$\text{error}(T, h, m) = \text{realization}(T, m) - \text{forecast}(T, h).$$

Tables 2 to 7 are organized along these lines. Table 2 shows recent forecasts from the SPF. Each column gives the projection for a different horizon or forecast step (h), beginning with that for the current quarter, defined as the quarter in which we conducted the survey. The dates (T) given in the rows show the periods forecast. These also correspond to the dates that we conducted the survey. Tables 3 to 6 report the recent projections of the four benchmark models. They are organized in the same way as Table 2. Table 7 reports recent values of the five alternative realizations (m) we use to compute the error statistics.

2.4. Qualifications.

We note two minor qualifications to the methods discussed above. The first concerns the vintage of data that we used to estimate and forecast the benchmark models for CPI inflation. The second concerns the five measures of realizations used for the unemployment rate, nonfarm payroll employment, and CPI inflation. To estimate and forecast the benchmark models for CPI inflation, we use the vintage of data that would have been available in the middle of each quarter. This postdates by one month the vintage that SPF panelists would have had at their disposal when they formed their projections.

To compute the realizations for unemployment, nonfarm payroll employment, and CPI inflation, we use the vintages associated with the middle of each quarter. The measure that we call initial comes from this vintage, even though the initial estimate was available in the vintage dated one month earlier. Thus, for these variables, our initial estimate reflects some revision by government statistical agencies. The effect for unemployment and CPI inflation is likely small. The effect could be somewhat larger for nonfarm payroll employment.

3. DESCRIPTION OF GRAPHS.

3.1. Root-Mean-Square Errors.

For each sample period shown in Table 1, we provide graphs of the root-mean-square error for the SPF forecast. There is one page for each sample period. On each page, we plot (for each forecast horizon) the RMSE on the y-axis. The x-axis shows the measure of the historical value that we used to compute the RMSE. These range from the value on its initial release to the value one quarter later to the value as we know it now (at the time we made the computation).

The graphs provide a tremendous amount of information. If we focus on a particular graph, we can see how a change in the measure of the realization (x-axis) affects the root-mean-square-error measure of accuracy. The effect is pronounced for some variables, such as real GDP and some of its components. For others, there is little or no effect. For example, because the historical data on interest rates are not revised, the estimated RMSE is the same in each case.

If we compare a particular point on one graph with the same point on another, we see how the forecast horizon affects accuracy. In general, the RMSE rises (accuracy falls) as the forecast horizon lengthens. Finally, if we compare a graph on one page with the corresponding graph on another page, we see how our estimates of accuracy in the SPF change with the sample period. Periods characterized by a high degree of economic turbulence will generally produce large RMSEs.

3.2. Fan Charts.

The last chart plots recent historical values and the latest SPF forecast. It also shows confidence intervals for the forecast, based on back-of-the-envelope calculations. The historical values and the SPF forecast are those associated with the latest vintage of data and survey, respectively, available at the time we ran our computer programs. The confidence intervals are constructed under the assumption that the historical forecast errors over the sample (shown in the footnote) follow a normal distribution with a mean of zero and a variance given by the squared root-mean-square error. The latter is estimated over the aforementioned sample, using the measure of history listed in the footnote.

Table 1A.
Forecast Error Summary Statistics for SPF Variable: CPI (CPI Inflation Rate)

Computed Over Various Sample Periods
Various Measures of Realizations
Transformation: Level
Lag Length for IAR(p), DAR(p), and DARM(p) Models: AIC

Source for Historical Realizations: Bureau of Labor Statistics via Haver Analytics

Last Updated: 02/23/2021 14:24

History: Initial Release

H	ME(S)	MAE(S)	RMSE(S)	RMSE(S/NC)	RMSE(S/IAR)	RMSE(S/DAR)	RMSE(S/DARM)	Nspf	N
1997:01-2018:03									
1	0.05	0.98	1.32	0.48	0.53	0.53	NA	87	87
2	0.02	1.43	2.14	0.67	0.86	0.88	NA	87	87
3	-0.05	1.48	2.17	0.69	0.91	0.90	NA	87	87
4	-0.11	1.48	2.15	0.70	0.92	0.93	NA	87	87
5	-0.15	1.50	2.15	0.76	0.92	0.94	NA	87	87

History: One Qtr After Initial Release

H	ME(S)	MAE(S)	RMSE(S)	RMSE(S/NC)	RMSE(S/IAR)	RMSE(S/DAR)	RMSE(S/DARM)	Nspf	N
1997:01-2018:03									
1	0.07	0.94	1.25	0.47	0.51	0.51	NA	87	87
2	0.04	1.39	2.06	0.66	0.85	0.86	NA	87	87
3	-0.03	1.45	2.10	0.68	0.90	0.88	NA	87	87
4	-0.09	1.45	2.09	0.69	0.92	0.92	NA	87	87
5	-0.13	1.47	2.09	0.75	0.91	0.94	NA	87	87

History: Five Qtrs After Initial Release

H	ME(S)	MAE(S)	RMSE(S)	RMSE(S/NC)	RMSE(S/IAR)	RMSE(S/DAR)	RMSE(S/DARM)	Nspf	N
1997:01-2018:03									
1	0.07	0.87	1.23	0.47	0.51	0.51	NA	87	87
2	0.03	1.35	2.04	0.66	0.84	0.86	NA	87	87
3	-0.03	1.39	2.07	0.68	0.90	0.88	NA	87	87
4	-0.10	1.39	2.06	0.68	0.92	0.92	NA	87	87
5	-0.14	1.40	2.05	0.75	0.91	0.94	NA	87	87

History: Nine Qtrs After Initial Release

H	ME(S)	MAE(S)	RMSE(S)	RMSE(S/NC)	RMSE(S/IAR)	RMSE(S/DAR)	RMSE(S/DARM)	Nspf	N
1997:01-2018:03									
1	0.07	0.84	1.21	0.47	0.51	0.51	NA	87	87
2	0.03	1.34	2.02	0.65	0.84	0.86	NA	87	87
3	-0.03	1.39	2.06	0.68	0.90	0.88	NA	87	87
4	-0.10	1.37	2.04	0.68	0.91	0.92	NA	87	87
5	-0.14	1.39	2.04	0.75	0.91	0.94	NA	87	87

History: Latest Vintage

1997:01-2018:03

H	ME(S)	MAE(S)	RMSE(S)	RMSE(S/NC)	RMSE(S/IAR)	RMSE(S/DAR)	RMSE(S/DARM)	Nspf	N
1	0.06	0.87	1.25	0.48	0.52	0.52	NA	87	87
2	0.02	1.38	2.05	0.65	0.85	0.87	NA	87	87
3	-0.04	1.43	2.09	0.69	0.90	0.89	NA	87	87
4	-0.11	1.40	2.07	0.69	0.91	0.92	NA	87	87
5	-0.15	1.41	2.06	0.75	0.91	0.94	NA	87	87

Notes for Table 1A.

(1) The forecast horizon is given by H, where H = 1 is the SPF forecast for the current quarter.
(2) The headers ME(S), MAE(S), and RMSE(S) are mean error, mean absolute error, and root-mean-square error for the SPF.
(3) The header RMSE(S/NC) is the ratio of the SPF RMSE to that of the no-change (NC) model.
(4) The headers RMSE(S/IAR), RMSE(S/DAR) and RMSE(S/DARM) are the ratios of the SPF RMSE to the RMSE of the iterated and direct autoregressive models and the direct autoregressive model augmented with monthly observations, respectively. All models are estimated on a rolling window of 60 observations from the Philadelphia Fed real-time data set.
(5) The headers Nspf and N are the number of observations analyzed for the SPF and benchmark models.
(6) When the variable forecast is a growth rate or an interest rate, it is expressed in annualized percentage points. When the variable forecast is the unemployment rate, it is expressed in percentage points.
(7) Sample periods refer to the dates forecast, not the dates when the forecasts were made.

Source: Tom Stark, Research Department, FRB Philadelphia.

Table 1B.
Ratios of Root-Mean-Square Errors for SPF Variable: CPI (CPI Inflation Rate)
Alternative P-Values in Parentheses

Computed Over Various Sample Periods
Various Measures of Realizations
Transformation: Level
Lag Length for IAR(p), DAR(p), and DARM(p) Models: AIC

Source for Historical Realizations: Bureau of Labor Statistics via Haver Analytics

Last Updated: 02/23/2021 14:24

History: Initial Release
1997:01-2018:03

H	RMSE(S/NC)	RMSE(S/IAR)	RMSE(S/DAR)	RMSE(S/DARM)	N1	N2	N3	N4
1	0.483	0.533	0.533	NA	87	87	87	NA
	(0.023)	(0.012)	(0.012)	(NA)				
	(0.026)	(0.015)	(0.015)	(NA)				
	(0.045)	(0.063)	(0.063)	(NA)				
2	0.668	0.865	0.878	NA	87	87	87	NA
	(0.028)	(0.051)	(0.091)	(NA)				
	(0.034)	(0.058)	(0.100)	(NA)				
	(0.031)	(0.103)	(0.147)	(NA)				
3	0.695	0.911	0.898	NA	87	87	87	NA
	(0.070)	(0.153)	(0.201)	(NA)				
	(0.082)	(0.169)	(0.218)	(NA)				
	(0.079)	(0.148)	(0.194)	(NA)				
4	0.697	0.923	0.933	NA	87	87	87	NA
	(0.041)	(0.062)	(0.078)	(NA)				
	(0.053)	(0.076)	(0.094)	(NA)				
	(0.066)	(0.054)	(0.048)	(NA)				
5	0.756	0.920	0.943	NA	87	87	87	NA
	(0.028)	(0.023)	(0.050)	(NA)				
	(0.041)	(0.034)	(0.066)	(NA)				
	(0.015)	(0.045)	(0.089)	(NA)				

History: One Qtr After Initial Release
1997:01-2018:03

H	RMSE(S/NC)	RMSE(S/IAR)	RMSE(S/DAR)	RMSE(S/DARM)	N1	N2	N3	N4
1	0.473	0.514	0.514	NA	87	87	87	NA
	(0.020)	(0.008)	(0.008)	(NA)				
	(0.023)	(0.010)	(0.010)	(NA)				
	(0.042)	(0.057)	(0.057)	(NA)				
2	0.659	0.847	0.859	NA	87	87	87	NA
	(0.030)	(0.055)	(0.089)	(NA)				
	(0.036)	(0.063)	(0.098)	(NA)				
	(0.032)	(0.104)	(0.144)	(NA)				
3	0.682	0.898	0.883	NA	87	87	87	NA
	(0.061)	(0.174)	(0.207)	(NA)				
	(0.073)	(0.190)	(0.224)	(NA)				
	(0.068)	(0.164)	(0.197)	(NA)				
4	0.695	0.919	0.922	NA	87	87	87	NA
	(0.034)	(0.073)	(0.090)	(NA)				
	(0.044)	(0.089)	(0.108)	(NA)				
	(0.057)	(0.062)	(0.060)	(NA)				
5	0.746	0.914	0.942	NA	87	87	87	NA
	(0.036)	(0.035)	(0.059)	(NA)				
	(0.050)	(0.049)	(0.077)	(NA)				
	(0.020)	(0.049)	(0.097)	(NA)				

History: Five Qtrs After Initial Release
1997:01-2018:03

H	RMSE(S/NC)	RMSE(S/IAR)	RMSE(S/DAR)	RMSE(S/DARM)	N1	N2	N3	N4
1	0.468	0.511	0.511	NA	87	87	87	NA
	(0.029)	(0.014)	(0.014)	(NA)				
	(0.033)	(0.017)	(0.017)	(NA)				
	(0.059)	(0.068)	(0.068)	(NA)				
2	0.658	0.843	0.856	NA	87	87	87	NA
	(0.040)	(0.053)	(0.077)	(NA)				
	(0.046)	(0.060)	(0.086)	(NA)				
	(0.045)	(0.100)	(0.128)	(NA)				
3	0.682	0.897	0.881	NA	87	87	87	NA
	(0.065)	(0.172)	(0.203)	(NA)				
	(0.077)	(0.188)	(0.220)	(NA)				
	(0.075)	(0.162)	(0.196)	(NA)				
4	0.681	0.917	0.922	NA	87	87	87	NA
	(0.025)	(0.074)	(0.107)	(NA)				
	(0.035)	(0.090)	(0.126)	(NA)				
	(0.049)	(0.071)	(0.072)	(NA)				
5	0.746	0.910	0.939	NA	87	87	87	NA
	(0.038)	(0.037)	(0.059)	(NA)				
	(0.052)	(0.052)	(0.077)	(NA)				
	(0.022)	(0.055)	(0.104)	(NA)				

History: Nine Qtrs After Initial Release
1997:01-2018:03

H	RMSE(S/NC)	RMSE(S/IAR)	RMSE(S/DAR)	RMSE(S/DARM)	N1	N2	N3	N4
1	0.468	0.512	0.512	NA	87	87	87	NA
	(0.034)	(0.017)	(0.017)	(NA)				
	(0.038)	(0.019)	(0.019)	(NA)				
	(0.063)	(0.077)	(0.077)	(NA)				
2	0.651	0.845	0.860	NA	87	87	87	NA
	(0.037)	(0.058)	(0.088)	(NA)				
	(0.044)	(0.066)	(0.097)	(NA)				
	(0.042)	(0.109)	(0.141)	(NA)				
3	0.679	0.899	0.883	NA	87	87	87	NA
	(0.061)	(0.186)	(0.202)	(NA)				
	(0.072)	(0.202)	(0.218)	(NA)				
	(0.069)	(0.177)	(0.193)	(NA)				
4	0.684	0.913	0.920	NA	87	87	87	NA
	(0.033)	(0.092)	(0.105)	(NA)				
	(0.043)	(0.109)	(0.123)	(NA)				
	(0.057)	(0.088)	(0.068)	(NA)				
5	0.753	0.909	0.939	NA	87	87	87	NA
	(0.037)	(0.039)	(0.070)	(NA)				
	(0.051)	(0.054)	(0.089)	(NA)				
	(0.020)	(0.059)	(0.120)	(NA)				

History: Latest Vintage
1997:01-2018:03

H	RMSE(S/NC)	RMSE(S/IAR)	RMSE(S/DAR)	RMSE(S/DARM)	N1	N2	N3	N4
1	0.477	0.521	0.521	NA	87	87	87	NA
	(0.028)	(0.014)	(0.014)	(NA)				
	(0.032)	(0.017)	(0.017)	(NA)				
	(0.054)	(0.073)	(0.073)	(NA)				
2	0.654	0.848	0.866	NA	87	87	87	NA
	(0.043)	(0.058)	(0.091)	(NA)				
	(0.050)	(0.066)	(0.101)	(NA)				
	(0.047)	(0.111)	(0.147)	(NA)				
3	0.687	0.902	0.889	NA	87	87	87	NA
	(0.065)	(0.189)	(0.209)	(NA)				
	(0.077)	(0.205)	(0.226)	(NA)				
	(0.075)	(0.183)	(0.202)	(NA)				
4	0.686	0.914	0.920	NA	87	87	87	NA
	(0.031)	(0.098)	(0.094)	(NA)				
	(0.042)	(0.116)	(0.111)	(NA)				
	(0.054)	(0.091)	(0.057)	(NA)				
5	0.751	0.908	0.939	NA	87	87	87	NA
	(0.023)	(0.045)	(0.064)	(NA)				
	(0.034)	(0.061)	(0.083)	(NA)				
	(0.011)	(0.065)	(0.111)	(NA)				

Notes for Table 1B.

(1) The forecast horizon is given by H, where H = 1 is the SPF forecast for the current quarter.
(2) The headers RMSE(S/NC), RMSE(S/IAR), RMSE(S/DAR), and RMSE(S/DARM) are the ratios of the SPF root-mean-square error to that of the benchmark models: No-change (NC), indirect autoregression (IAR), direct autoregession (DAR), and direct autoregression augmented with monthly information (DARM). These statistics may differ slightly from those reported in Table 1A because they incorporate only those observations common to both the SPF and the benchmark model. The previous statistics make use of all available observations for each model.
(3) All models are estimated on a rolling window of 60 observations from the Philadelphia Fed real-time data set.
(4) A set of three two-sided p-values (in parentheses) accompanies each statistic. These are the p-values for the test of the equality of mean-square-error. The first is for the Diebold-Mariano (1995, JBES) statistic, using a uniform lag window with the trunction lag set to the forecast horizon minus one. (The tables report the p-values using a Bartlett window when the uniform window produces a negative standard error.) The second is for the Harvey-Leybourne-Newbold (1997, IJF) correction to the Diebold-Mariano statistic. The third is for the Diebold-Mariano statistic, using a Bartlett lag window with the truncation lag increased four quarters.
(5) The headers N1, N2, N3, and N4 show the number of observations used in constructing each ratio of root-mean-square errors.
(6) Sample periods refer to the dates forecast, not the dates when the forecasts were made.

Source: Tom Stark, Research Department, FRB Philadelphia.

Table 2. Recent SPF Forecasts
(Dated at the Quarter Forecast)

Variable: CPI (CPI Inflation Rate)
By Forecast Step (1 to 5)
Transformation: Level

Last Updated: 02/23/2021 14:24

Qtr Forecast	Step 1	Step 2	Step 3	Step 4	Step 5
2014:02	1.879	1.655	1.995	1.900	2.166
2014:03	2.200	1.923	1.935	2.001	2.100
2014:04	1.040	2.036	1.900	1.963	2.140
2015:01	-1.385	1.800	2.100	2.010	2.000
2015:02	1.900	1.620	1.900	2.160	2.000
2015:03	2.000	1.950	1.900	1.965	2.100
2015:04	0.886	1.799	1.900	1.956	2.000
2016:01	0.400	1.802	2.000	1.942	2.100
2016:02	1.900	1.600	2.138	2.157	2.100
2016:03	2.000	2.003	2.115	2.064	2.213
2016:04	2.207	2.200	2.100	2.100	2.171
2017:01	2.490	2.176	2.200	2.010	2.125
2017:02	1.560	2.288	2.200	2.210	2.117
2017:03	1.647	2.200	2.300	2.220	2.300
2017:04	2.300	2.318	2.300	2.450	2.215
2018:01	2.700	2.070	2.200	2.400	2.400
2018:02	1.906	2.020	2.018	2.123	2.200
2018:03	2.300	2.201	2.300	2.175	2.179
2018:04	2.336	2.300	2.290	2.227	2.100
2019:01	1.100	2.413	2.376	2.300	2.300
2019:02	2.406	2.300	2.300	2.061	2.248
2019:03	1.880	2.124	2.307	2.350	2.300
2019:04	1.800	2.088	2.100	2.200	2.400
2020:01	2.000	2.200	2.005	2.118	2.253
2020:02	-2.579	2.000	2.100	2.000	2.000
2020:03	2.337	1.500	2.200	2.154	2.054
2020:04	2.008	1.612	1.937	2.233	2.100
2021:01	2.496	2.000	1.768	2.000	2.200
2021:02	NA	2.132	2.000	1.600	2.000
2021:03	NA	NA	2.086	2.100	2.100
2021:04	NA	NA	NA	2.200	2.168
2022:01	NA	NA	NA	NA	2.200

Notes for Table 2.

(1) Each column gives the sequence of SPF projections for a given forecast step. The forecast steps range from one (the forecast for the quarter in which the survey was conducted) to four quarters in the future (step 5).
(2) The dates listed in the rows are the dates forecast, not the dates when the forecasts were made, with the exception of the forecast at step one, for which the two dates coincide.

Source: Tom Stark, Research Department, FRB Philadelphia.

Table 3. Recent Benchmark Model 1 IAR Forecasts
(Dated at the Quarter Forecast)

Variable: CPI (CPI Inflation Rate)
By Forecast Step (1 to 5)
Transformation: Level
Lag Length for IAR(p): AIC

Source for Historical Realizations: Bureau of Labor Statistics via Haver Analytics

Last Updated: 02/23/2021 14:24

Qtr Forecast	Step 1	Step 2	Step 3	Step 4	Step 5
2014:02	2.335	2.375	2.441	2.421	2.453
2014:03	2.541	2.413	2.411	2.440	2.423
2014:04	2.155	2.450	2.428	2.418	2.440
2015:01	1.748	2.348	2.433	2.430	2.419
2015:02	0.929	2.825	2.383	2.430	2.431
2015:03	2.365	1.879	2.546	2.389	2.429
2015:04	2.050	2.240	2.105	2.291	2.390
2016:01	1.742	2.142	2.215	2.159	2.283
2016:02	1.555	2.043	2.161	2.210	2.172
2016:03	2.152	1.942	2.104	2.164	2.209
2016:04	1.980	2.075	2.023	2.116	2.165
2017:01	2.319	2.050	2.060	2.039	2.118
2017:02	2.360	2.173	2.064	2.057	2.043
2017:03	1.610	2.201	2.144	2.067	2.056
2017:04	2.077	1.988	2.169	2.138	2.067
2018:01	2.346	2.089	2.063	2.163	2.137
2018:02	2.225	2.160	2.091	2.077	2.162
2018:03	1.809	1.905	2.124	2.092	2.080
2018:04	2.162	2.141	2.034	2.117	2.092
2019:01	1.976	2.147	2.198	2.113	2.116
2019:02	1.860	2.207	2.114	2.158	2.111
2019:03	2.544	2.250	2.178	2.108	2.140
2019:04	1.829	2.050	2.168	2.127	2.112
2020:01	2.165	2.054	1.989	2.070	2.119
2020:02	1.739	2.001	2.107	2.067	2.059
2020:03	0.513	2.108	1.999	2.078	2.100
2020:04	3.691	2.674	2.103	2.030	2.060
2021:01	1.459	1.609	2.426	2.030	2.039
2021:02	NA	1.700	1.492	1.894	2.011
2021:03	NA	NA	1.922	1.897	1.795
2021:04	NA	NA	NA	1.917	2.002
2022:01	NA	NA	NA	NA	1.875

Notes for Table 3.

(1) Each column gives the sequence of benchmark IAR projections for a given forecast step. The forecast steps range from one to five. The first step corresponds to the forecast that SPF panelists make for the quarter in which the survey is conducted.
(2) The dates listed in the rows are the dates forecast, not the dates when the forecasts were made, with the exception of the forecast at step one, for which the two dates coincide.
(3) The IAR benchmark model is estimated on a fixed 60-quarter rolling window. Its forecasts are computed with the indirect method. Estimation uses data from the Philadelphia Fed real-time data set.

Source: Tom Stark, Research Department, FRB Philadelphia.

Table 4. Recent Benchmark Model 2 No-Change Forecasts
(Dated at the Quarter Forecast)

Variable: CPI (CPI Inflation Rate)
By Forecast Step (1 to 5)
Transformation: Level

Source for Historical Realizations: Bureau of Labor Statistics via Haver Analytics

Last Updated: 02/23/2021 14:24

Qtr Forecast	Step 1	Step 2	Step 3	Step 4	Step 5
2014:02	1.912	1.136	2.629	-0.028	1.436
2014:03	3.031	1.912	1.136	2.629	-0.028
2014:04	1.096	3.031	1.912	1.136	2.629
2015:01	-1.199	1.096	3.031	1.912	1.136
2015:02	-3.059	-1.199	1.096	3.031	1.912
2015:03	2.978	-3.059	-1.199	1.096	3.031
2015:04	1.583	2.978	-3.059	-1.199	1.096
2016:01	0.227	1.583	2.978	-3.059	-1.199
2016:02	-0.312	0.227	1.583	2.978	-3.059
2016:03	2.542	-0.312	0.227	1.583	2.978
2016:04	1.632	2.542	-0.312	0.227	1.583
2017:01	3.040	1.632	2.542	-0.312	0.227
2017:02	3.147	3.040	1.632	2.542	-0.312
2017:03	-0.312	3.147	3.040	1.632	2.542
2017:04	2.014	-0.312	3.147	3.040	1.632
2018:01	3.309	2.014	-0.312	3.147	3.040
2018:02	3.508	3.309	2.014	-0.312	3.147
2018:03	1.656	3.508	3.309	2.014	-0.312
2018:04	1.996	1.656	3.508	3.309	2.014
2019:01	1.486	1.996	1.656	3.508	3.309
2019:02	0.877	1.486	1.996	1.656	3.508
2019:03	2.918	0.877	1.486	1.996	1.656
2019:04	1.789	2.918	0.877	1.486	1.996
2020:01	2.374	1.789	2.918	0.877	1.486
2020:02	1.208	2.374	1.789	2.918	0.877
2020:03	-3.530	1.208	2.374	1.789	2.918
2020:04	5.158	-3.530	1.208	2.374	1.789
2021:01	2.430	5.158	-3.530	1.208	2.374
2021:02	NA	2.430	5.158	-3.530	1.208
2021:03	NA	NA	2.430	5.158	-3.530
2021:04	NA	NA	NA	2.430	5.158
2022:01	NA	NA	NA	NA	2.430

Notes for Table 4.

(1) Each column gives the sequence of benchmark no-change projections for a given forecast step. The forecast steps range from one to five. The first step corresponds to the forecast that SPF panelists make for the quarter in which the survey is conducted.
(2) The dates listed in the rows are the dates forecast, not the dates when the forecasts were made, with the exception of the forecast at step one, for which the two dates coincide.
(3) The projections use data from the Philadelphia Fed real-time data set.

Source: Tom Stark, Research Department, FRB Philadelphia.

Table 5. Recent Benchmark Model 3 DAR Forecasts
(Dated at the Quarter Forecast)

Variable: CPI (CPI Inflation Rate)
By Forecast Step (1 to 5)
Transformation: Level
Lag Length for DAR(p): AIC

Source for Historical Realizations: Bureau of Labor Statistics via Haver Analytics

Last Updated: 02/23/2021 14:24

Qtr Forecast	Step 1	Step 2	Step 3	Step 4	Step 5
2014:02	2.335	2.591	2.420	2.816	2.404
2014:03	2.541	2.498	2.540	2.408	2.309
2014:04	2.155	2.349	2.476	2.632	2.447
2015:01	1.748	2.568	2.376	2.518	2.352
2015:02	0.929	2.823	2.506	2.331	2.401
2015:03	2.365	2.827	2.597	2.616	2.461
2015:04	2.050	2.132	2.665	2.857	2.311
2016:01	1.742	2.242	2.158	2.947	2.079
2016:02	1.555	2.357	2.234	2.134	1.729
2016:03	2.152	2.323	2.226	2.284	2.258
2016:04	1.980	2.013	2.212	2.368	2.100
2017:01	2.319	2.108	2.033	2.280	1.926
2017:02	2.360	2.027	2.095	2.028	1.781
2017:03	1.610	2.044	2.071	2.110	2.086
2017:04	2.077	2.343	2.099	2.041	2.013
2018:01	2.346	2.107	2.202	2.056	2.216
2018:02	2.225	1.975	2.097	2.319	2.258
2018:03	1.809	1.941	2.065	2.098	1.831
2018:04	2.162	2.191	2.023	2.012	2.088
2019:01	1.976	2.135	2.164	1.999	2.246
2019:02	1.860	2.208	2.129	2.167	2.278
2019:03	2.544	2.244	2.158	2.129	2.081
2019:04	1.829	1.976	2.138	2.166	2.111
2020:01	2.165	2.103	2.040	2.178	2.054
2020:02	1.739	1.993	2.080	2.023	1.946
2020:03	0.513	2.137	2.017	2.089	2.166
2020:04	3.691	2.749	2.070	2.016	2.025
2021:01	1.459	1.401	2.222	2.078	2.060
2021:02	NA	1.792	1.636	2.395	1.919
2021:03	NA	NA	1.839	1.604	1.204
2021:04	NA	NA	NA	2.198	2.273
2022:01	NA	NA	NA	NA	1.914

Notes for Table 5.

(1) Each column gives the sequence of benchmark DAR projections for a given forecast step. The forecast steps range from one to five. The first step corresponds to the forecast that SPF panelists make for the quarter in which the survey is conducted.
(2) The dates listed in the rows are the dates forecast, not the dates when the forecasts were made, with the exception of the forecast at step one, for which the two dates coincide.
(3) The DAR benchmark model is estimated on a fixed 60-quarter rolling window. Its forecasts are computed with the direct method. Estimation uses data from the Philadelphia Fed real-time data set.

Source: Tom Stark, Research Department, FRB Philadelphia.

Table 6. Recent Benchmark Model 4 DARM Forecasts
(Dated at the Quarter Forecast)

Variable: CPI (CPI Inflation Rate)
By Forecast Step (1 to 5)
Transformation: Level
Lag Length for DARM(p): AIC

Source for Historical Realizations: Bureau of Labor Statistics via Haver Analytics

Last Updated: 02/23/2021 14:24

Qtr Forecast	Step 1	Step 2	Step 3	Step 4	Step 5
2014:02	NA	NA	NA	NA	NA
2014:03	NA	NA	NA	NA	NA
2014:04	NA	NA	NA	NA	NA
2015:01	NA	NA	NA	NA	NA
2015:02	NA	NA	NA	NA	NA
2015:03	NA	NA	NA	NA	NA
2015:04	NA	NA	NA	NA	NA
2016:01	NA	NA	NA	NA	NA
2016:02	NA	NA	NA	NA	NA
2016:03	NA	NA	NA	NA	NA
2016:04	NA	NA	NA	NA	NA
2017:01	NA	NA	NA	NA	NA
2017:02	NA	NA	NA	NA	NA
2017:03	NA	NA	NA	NA	NA
2017:04	NA	NA	NA	NA	NA
2018:01	NA	NA	NA	NA	NA
2018:02	NA	NA	NA	NA	NA
2018:03	NA	NA	NA	NA	NA
2018:04	NA	NA	NA	NA	NA
2019:01	NA	NA	NA	NA	NA
2019:02	NA	NA	NA	NA	NA
2019:03	NA	NA	NA	NA	NA
2019:04	NA	NA	NA	NA	NA
2020:01	NA	NA	NA	NA	NA
2020:02	NA	NA	NA	NA	NA
2020:03	NA	NA	NA	NA	NA
2020:04	NA	NA	NA	NA	NA
2021:01	NA	NA	NA	NA	NA
2021:02	NA	NA	NA	NA	NA
2021:03	NA	NA	NA	NA	NA
2021:04	NA	NA	NA	NA	NA
2022:01	NA	NA	NA	NA	NA

Notes for Table 6.

(1) Each column gives the sequence of benchmark DARM projections for a given forecast step. The forecast steps range from one to five. The first step corresponds to the forecast that SPF panelists make for the quarter in which the survey is conducted.
(2) The dates listed in the rows are the dates forecast, not the dates when the forecasts were made, with the exception of the forecast at step one, for which the two dates coincide.
(3) The DARM benchmark model is estimated on a fixed 60-quarter rolling window. Its forecasts are computed with the direct method and incorporate recent monthly values of the dependent variable. Estimation uses data from the Philadelphia Fed real-time data set.

Source: Tom Stark, Research Department, FRB Philadelphia.

Table 7. Recent Realizations (Various Measures)
Philadelphia Fed Real-Time Data Set

Variable: CPI (CPI Inflation Rate)
Transformation: Level

Source for Historical Realizations: Bureau of Labor Statistics via Haver Analytics

Last Updated: 02/23/2021 14:24

Column (1): Initial Release
Column (2): One Qtr After Initial Release
Column (3): Five Qtrs After Initial Release
Column (4): Nine Qtrs After Initial Release
Column (5): Latest Vintage

Obs. Date	(1)	(2)	(3)	(4)	(5)
2014:02	3.031	3.031	2.437	1.915	2.141
2014:03	1.096	1.096	1.183	1.037	1.027
2014:04	-1.199	-0.854	-0.312	-0.701	-0.990
2015:01	-3.059	-3.059	-2.863	-2.521	-2.564
2015:02	2.978	2.978	2.439	2.353	2.756
2015:03	1.583	1.583	1.496	1.536	1.519
2015:04	0.227	0.769	0.358	0.120	-0.030
2016:01	-0.312	-0.312	0.113	-0.067	-0.248
2016:02	2.542	2.526	2.328	2.725	3.238
2016:03	1.632	1.784	1.843	1.875	1.709
2016:04	3.040	3.040	2.737	2.624	2.563
2017:01	3.147	3.147	2.958	2.780	2.700
2017:02	-0.312	-0.312	0.100	0.386	0.715
2017:03	2.014	2.126	2.153	2.156	1.911
2017:04	3.309	3.309	3.142	3.119	3.153
2018:01	3.508	3.508	3.237	3.250	3.146
2018:02	1.656	1.656	2.149	2.195	2.525
2018:03	1.996	2.010	2.078	1.632	1.632
2018:04	1.486	1.486	1.300	NA	1.572
2019:01	0.877	0.877	0.918	NA	0.710
2019:02	2.918	2.918	3.027	NA	3.501
2019:03	1.789	1.821	1.288	NA	1.288
2019:04	2.374	2.374	NA	NA	2.630
2020:01	1.208	1.208	NA	NA	0.997
2020:02	-3.530	-3.530	NA	NA	-3.101
2020:03	5.158	4.680	NA	NA	4.680
2020:04	2.430	NA	NA	NA	2.430
2021:01	NA	NA	NA	NA	NA
2021:02	NA	NA	NA	NA	NA
2021:03	NA	NA	NA	NA	NA
2021:04	NA	NA	NA	NA	NA
2022:01	NA	NA	NA	NA	NA

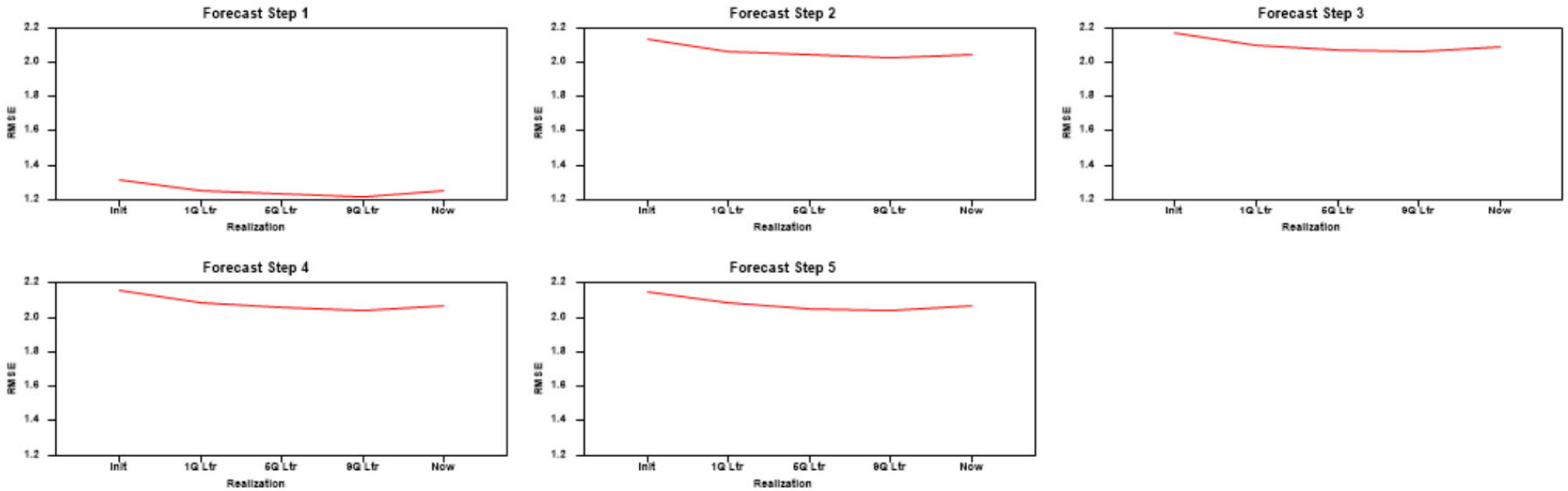
Notes for Table 7.

(1) Each column reports a sequence of realizations from the Philadelphia Fed real-time data set.
(2) The date listed in each row is the observation date.
(3) Moving across a particular row shows how the observation is revised in subsequent releases.

Source: Tom Stark, Research Department, FRB Philadelphia.

Root-Mean-Square Errors: 1997:01-2018:03

SPF Projections for CPI Inflation Rate, Transformation: Level



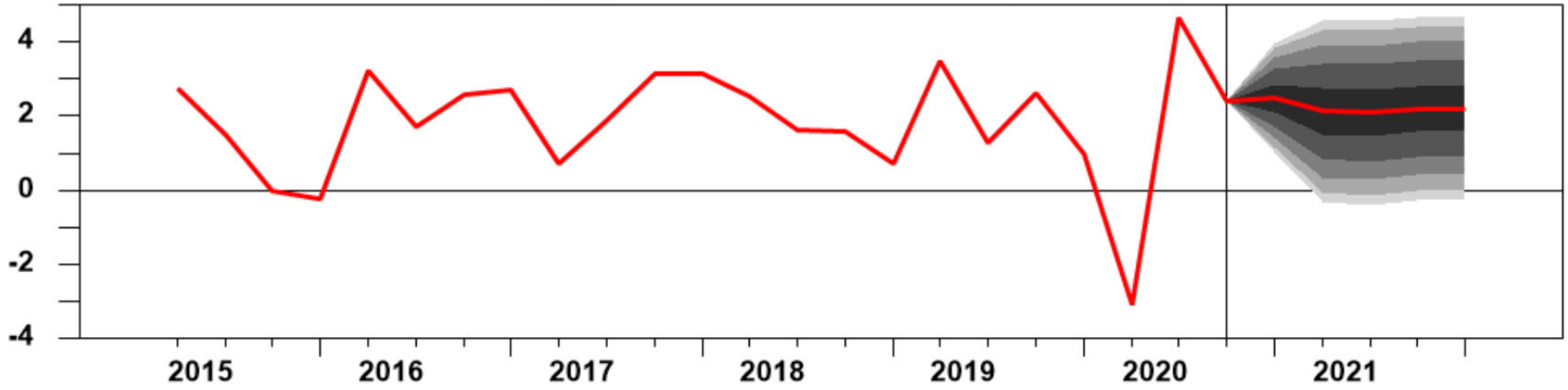
The RMSE is plotted against the realization used to compute it, from the value on initial release to the value as we now know it. Source: Tom Stark, FRB Philadelphia.

CPI Inflation Rate

History, Forecasts, and Ranges for the SPF of 2021:01

Level

(Ranges Cover 25 to 80 Percent Confidence)



Ranges at each horizon use the N(0,MSE) density. The MSEs are based on the sample 94:03-19:03 and use the realization: Five Qtrs After Initial Release. Source: T.Stark, FRB Phila.