Error Statistics for the Survey of Professional Forecasters:
Nonfarm Payroll Employment

Release Date: 05/24/2024

```
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

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

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.

4. SPECIAL TREATMENT OF COVID-19 EXTREME HISTORICAL OBSERVATIONS.

Many macroeconomic variables experienced extreme values in 2020:Q2 and 2020:Q3 due to the partial shutdown of the U.S. economy at the end of March 2020. In some cases, these extreme values adversely affected the estimation of parameters in the benchmark models. The effect produced unrealistic parameter values in the models. In some cases, the models became dynamically unstable in sample periods that encompassed the 2020:Q2 and 2020:Q3 observations, leading to distorted forecast error statistics in those benchmark models for some survey variables. Comparisons with the SPF projections became hard to defend.

Beginning with the error statistics published following the 2022:Q1 survey, we proceed in the following way. For some survey variables, we scale back the magnitude of the historical extreme observations for 2020:Q2 and 2020:Q3. We make these adjustment before the estimation of parameters in the benchmark models but reverse the adjustment prior to forecasting the models. The intent is to prevent adverse, unrealistic effects on the parameter estimates while allowing the extreme historical observations to condition the projections. These adjustments to extreme historical observations do not change the historical realizations used to compute forecast errors or forecast error statistics: We continue to use unadjusted, official U.S. government data for these purposes.

The survey variables for which we make the adjustments are: Nominal GDP, Unemployment, Employment, Industrial Production, Real GDP, and Real Personal Consumption Expenditures. We do not adjust the historical values for the remaining variables.

Table 1A.
Forecast Error Summary Statistics for SPF Variable: EMP (Nonfarm Payroll Employment)

Computed Over Various Sample Periods
Various Measures of Realizations
Transformation: Q/Q Growth Rate
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: 05/24/2024 12:50

H	ME(S)	MAE(S)	RMSE(S)	RMSE(S/NC)	RMSE(S/IAR)	RMSE(S/DAR)	RMSE(S/DARM)	Nspf	N
			History: Initial Release						
2004:04-2021:04									
1	0.17	0.45	1.11	0.12	0.10	0.10	0.30	69	69
2	-0.70	1.13	5.05	0.60	0.54	0.56	0.52	69	69
3	-0.48	1.56	5.74	0.72	0.70	0.87	0.62	69	69
4	-0.48	1.69	5.81	0.71	0.77	0.92	0.65	69	69
5	-0.48	1.72	5.84	0.69	0.82	0.94	0.64	69	69
H	ME(S)	MAE(S)	RMSE(S)	RMSE(S/NC)	RMSE(S/IAR)	RMSE(S/DAR)	RMSE(S/DARM)	Nspf	N
			History: One Qtr After Initial Release						
2004:04-2021:04									
1	0.14	0.49	1.15	0.12	0.11	0.11	0.31	69	69
2	-0.73	1.18	5.06	0.61	0.54	0.56	0.52	69	69
3	-0.51	1.57	5.76	0.72	0.70	0.88	0.62	69	69
4	-0.51	1.70	5.83	0.70	0.77	0.92	0.65	69	69
5	-0.51	1.74	5.85	0.70	0.84	0.95	0.63	69	69
H	ME(S)	MAE(S)	RMSE(S)	RMSE(S/NC)	RMSE(S/IAR)	RMSE(S/DAR)	RMSE(S/DARM)	Nspf	N
			History: Five Qtrs After Initial Release						
2004:04-2021:04									
1	0.12	0.56	1.20	0.13	0.11	0.11	0.33	69	69
2	-0.75	1.20	5.08	0.61	0.55	0.57	0.53	69	69
3	-0.53	1.58	5.70	0.70	0.68	0.86	0.61	69	69
4	-0.53	1.72	5.77	0.71	0.77	0.92	0.64	69	69
5	-0.52	1.83	5.82	0.69	0.83	0.94	0.64	69	69
H	ME(S)	MAE(S)	RMSE(S)	RMSE(S/NC)	RMSE(S/IAR)	RMSE(S/DAR)	RMSE(S/DARM)	Nspf	N
			History: Nine Qtrs After Initial Release						
2004:04-2021:04									
1	0.12	0.55	1.20	0.13	0.11	0.11	0.33	69	69
2	-0.75	1.21	5.06	0.61	0.54	0.57	0.53	69	69
3	-0.53	1.59	5.71	0.71	0.69	0.86	0.61	69	69
4	-0.53	1.75	5.79	0.71	0.77	0.92	0.64	69	69
5	-0.53	1.82	5.83	0.69	0.83	0.94	0.64	69	69

H	ME(S)	MAE(S)	RMSE(S)	RMSE(S/NC)	RMSE(S/IAR)	RMSE(S/DAR)	RMSE(S/DARM)	Nspf	N

History: Latest Vintage

2004:04-2021:04

H	ME(S)	MAE(S)	RMSE(S)	RMSE(S/NC)	RMSE(S/IAR)	RMSE(S/DAR)	RMSE(S/DARM)	Nspf	N
1	0.12	0.55	1.21	0.13	0.11	0.11	0.33	69	69
2	-0.75	1.21	5.05	0.61	0.54	0.57	0.53	69	69
3	-0.53	1.61	5.72	0.71	0.69	0.87	0.62	69	69
4	-0.53	1.75	5.80	0.71	0.77	0.92	0.64	69	69
5	-0.52	1.82	5.83	0.69	0.83	0.94	0.64	69	69

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: EMP (Nonfarm Payroll Employment)
Alternative P-Values in Parentheses
```

```
Computed Over Various Sample Periods
Various Measures of Realizations
Transformation: Q/Q Growth Rate
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: 05/24/2024 12:50
```

History: Initial Release
2004:04-2021:04

H	RMSE(S/NC)	RMSE(S/IAR)	RMSE(S/DAR)	RMSE(S/DARM)	N1	N2	N3	N4
1	0.119	0.104	0.104	0.296	69	69	69	69
	(0.160)	(0.203)	(0.203)	(0.059)				
	(0.168)	(0.210)	(0.210)	(0.065)				
	(0.267)	(0.275)	(0.275)	(0.235)				
2	0.605	0.543	0.563	0.520	69	69	69	69
	(0.279)	(0.290)	(0.244)	(0.296)				
	(0.293)	(0.304)	(0.259)	(0.310)				
	(0.260)	(0.272)	(0.249)	(0.279)				
3	0.717	0.699	0.875	0.621	69	69	69	69
	(0.281)	(0.277)	(0.220)	(0.286)				
	(0.302)	(0.298)	(0.242)	(0.307)				
	(0.272)	(0.274)	(0.218)	(0.284)				
4	0.709	0.768	0.918	0.649	69	69	69	69
	(0.281)	(0.269)	(0.230)	(0.286)				
	(0.310)	(0.297)	(0.258)	(0.314)				
	(0.279)	(0.266)	(0.216)	(0.281)				
5	0.687	0.824	0.940	0.640	69	69	69	69
	(0.276)	(0.224)	(0.153)	(0.288)				
	(0.312)	(0.260)	(0.186)	(0.324)				
	(0.274)	(0.221)	(0.156)	(0.277)				

History: One Qtr After Initial Release
2004:04-2021:04

H	RMSE(S/NC)	RMSE(S/IAR)	RMSE(S/DAR)	RMSE(S/DARM)	N1	N2	N3	N4
1	0.123	0.107	0.107	0.307	69	69	69	69
	(0.161)	(0.204)	(0.204)	(0.063)				
	(0.169)	(0.212)	(0.212)	(0.069)				
	(0.268)	(0.276)	(0.276)	(0.237)				
2	0.606	0.544	0.564	0.519	69	69	69	69
	(0.282)	(0.293)	(0.247)	(0.299)				
	(0.296)	(0.307)	(0.261)	(0.313)				
	(0.262)	(0.273)	(0.250)	(0.280)				
3	0.718	0.700	0.876	0.623	69	69	69	69
	(0.281)	(0.277)	(0.219)	(0.287)				
	(0.302)	(0.298)	(0.240)	(0.309)				
	(0.272)	(0.274)	(0.218)	(0.284)				
4	0.705	0.767	0.915	0.651	69	69	69	69
	(0.282)	(0.269)	(0.231)	(0.285)				
	(0.311)	(0.297)	(0.260)	(0.314)				
	(0.278)	(0.265)	(0.215)	(0.281)				
5	0.700	0.837	0.948	0.633	69	69	69	69
	(0.274)	(0.215)	(0.125)	(0.289)				
	(0.310)	(0.251)	(0.156)	(0.325)				
	(0.272)	(0.211)	(0.128)	(0.277)				

History: Five Qtrs After Initial Release
2004:04-2021:04

H	RMSE(S/NC)	RMSE(S/IAR)	RMSE(S/DAR)	RMSE(S/DARM)	N1	N2	N3	N4
1	0.131	0.114	0.114	0.330	69	69	69	69
	(0.160)	(0.202)	(0.202)	(0.054)				
	(0.168)	(0.210)	(0.210)	(0.060)				
	(0.268)	(0.276)	(0.276)	(0.232)				
2	0.612	0.547	0.574	0.532	69	69	69	69
	(0.283)	(0.292)	(0.251)	(0.298)				
	(0.297)	(0.306)	(0.265)	(0.313)				
	(0.264)	(0.275)	(0.251)	(0.280)				
3	0.701	0.684	0.859	0.615	69	69	69	69
	(0.283)	(0.279)	(0.228)	(0.287)				
	(0.305)	(0.300)	(0.250)	(0.308)				
	(0.274)	(0.276)	(0.226)	(0.285)				
4	0.708	0.769	0.916	0.635	69	69	69	69
	(0.283)	(0.270)	(0.226)	(0.286)				
	(0.312)	(0.299)	(0.254)	(0.315)				
	(0.279)	(0.266)	(0.208)	(0.281)				
5	0.695	0.832	0.945	0.639	69	69	69	69
	(0.275)	(0.219)	(0.125)	(0.288)				
	(0.311)	(0.255)	(0.156)	(0.324)				
	(0.274)	(0.216)	(0.131)	(0.276)				

History: Nine Qtrs After Initial Release
2004:04-2021:04

H	RMSE(S/NC)	RMSE(S/IAR)	RMSE(S/DAR)	RMSE(S/DARM)	N1	N2	N3	N4
1	0.131	0.114	0.114	0.329	69	69	69	69
	(0.162)	(0.204)	(0.204)	(0.057)				
	(0.169)	(0.211)	(0.211)	(0.063)				
	(0.268)	(0.276)	(0.276)	(0.236)				
2	0.607	0.543	0.568	0.528	69	69	69	69
	(0.282)	(0.292)	(0.249)	(0.298)				
	(0.297)	(0.306)	(0.264)	(0.312)				
	(0.263)	(0.274)	(0.251)	(0.280)				
3	0.706	0.689	0.864	0.615	69	69	69	69
	(0.282)	(0.277)	(0.222)	(0.287)				
	(0.303)	(0.299)	(0.244)	(0.308)				
	(0.273)	(0.275)	(0.220)	(0.285)				
4	0.708	0.769	0.917	0.641	69	69	69	69
	(0.284)	(0.271)	(0.229)	(0.286)				
	(0.312)	(0.300)	(0.257)	(0.314)				
	(0.280)	(0.268)	(0.212)	(0.281)				
5	0.694	0.831	0.943	0.638	69	69	69	69
	(0.275)	(0.220)	(0.129)	(0.288)				
	(0.311)	(0.256)	(0.161)	(0.324)				
	(0.274)	(0.217)	(0.134)	(0.276)				

```
                    History: Latest Vintage
                    2004:04-2021:04

  H    RMSE(S/NC) RMSE(S/IAR) RMSE(S/DAR) RMSE(S/DARM) N1  N2  N3  N4
   1     0.131       0.114       0.114        0.328    69  69  69  69
        (0.162)     (0.204)     (0.204)      (0.059)
        (0.170)     (0.212)     (0.212)      (0.065)
        (0.268)     (0.276)     (0.276)      (0.237)

   2     0.606       0.543       0.566        0.525    69  69  69  69
        (0.282)     (0.292)     (0.249)      (0.298)
        (0.296)     (0.306)     (0.263)      (0.312)
        (0.263)     (0.274)     (0.251)      (0.280)

   3     0.709       0.692       0.867        0.616    69  69  69  69
        (0.282)     (0.277)     (0.221)      (0.287)
        (0.303)     (0.299)     (0.242)      (0.308)
        (0.273)     (0.275)     (0.219)      (0.285)

   4     0.708       0.769       0.917        0.643    69  69  69  69
        (0.284)     (0.272)     (0.231)      (0.286)
        (0.313)     (0.301)     (0.260)      (0.314)
        (0.281)     (0.269)     (0.215)      (0.281)

   5     0.694       0.830       0.943        0.637    69  69  69  69
        (0.275)     (0.221)     (0.130)      (0.288)
        (0.311)     (0.256)     (0.162)      (0.324)
        (0.274)     (0.218)     (0.134)      (0.277)
```

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: EMP (Nonfarm Payroll Employment)
By Forecast Step (1 to 5)
Transformation: Q/Q Growth Rate

Last Updated: 05/24/2024 12:50

_____
```

Qtr Forecast	Step 1	Step 2	Step 3	Step 4	Step 5
2017:03	1.538	1.408	1.394	1.371	1.317
2017:04	1.506	1.323	1.358	1.318	1.366
2018:01	1.503	1.352	1.274	1.366	1.291
2018:02	1.475	1.397	1.364	1.327	1.300
2018:03	1.600	1.420	1.373	1.278	1.346
2018:04	1.642	1.400	1.297	1.230	1.262
2019:01	1.600	1.387	1.300	1.297	1.398
2019:02	1.526	1.255	1.348	1.300	1.218
2019:03	1.201	1.313	1.345	1.275	1.200
2019:04	1.296	1.178	1.230	1.130	1.137
2020:01	1.412	1.055	1.090	1.055	1.138
2020:02	-48.012	1.332	1.260	1.152	1.218
2020:03	19.899	23.463	1.044	0.964	0.853
2020:04	6.011	3.462	8.182	0.915	1.000
2021:01	1.210	4.021	3.127	4.526	0.895
2021:02	4.862	3.365	3.572	1.703	6.470
2021:03	5.876	6.376	3.761	3.715	2.425
2021:04	3.876	4.215	3.987	4.747	3.305
2022:01	3.565	3.329	3.839	3.037	3.646
2022:02	3.077	2.886	2.509	3.277	2.317
2022:03	2.738	2.346	2.057	2.580	2.119
2022:04	1.721	1.326	1.530	1.793	2.440
2023:01	2.222	0.619	0.699	1.252	1.144
2023:02	1.211	0.004	0.280	0.758	0.480
2023:03	1.304	0.338	0.370	0.327	0.632
2023:04	1.145	0.798	0.197	0.488	-0.113
2024:01	1.814	0.503	0.434	0.286	0.473
2024:02	1.529	0.915	0.749	0.598	0.175
2024:03	NA	1.120	0.873	0.622	0.596
2024:04	NA	NA	0.983	0.930	0.907
2025:01	NA	NA	NA	1.091	1.013
2025:02	NA	NA	NA	NA	0.820

```
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: EMP (Nonfarm Payroll Employment)
By Forecast Step (1 to 5)
Transformation: Q/Q Growth Rate
Lag Length for IAR(p):  AIC

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

Last Updated: 05/24/2024 12:50

_____

Qtr Forecast Step 1  Step 2  Step 3  Step 4  Step 5
2017:03       1.263   1.229   1.064   1.144   0.604
2017:04       1.168   1.137   1.104   0.974   1.022
2018:01       1.453   1.083   1.044   1.024   0.928
2018:02       1.605   1.344   0.999   0.962   0.956
2018:03       1.686   1.526   1.235   0.939   0.903
2018:04       1.556   1.567   1.392   1.133   0.892
2019:01       1.578   1.426   1.436   1.281   1.049
2019:02       1.539   1.433   1.296   1.316   1.178
2019:03       1.045   1.405   1.289   1.188   1.216
2019:04       1.205   0.881   1.274   1.175   1.102
2020:01       1.622   1.195   0.822   1.168   1.084
2020:02       0.676   1.542   1.121   0.778   1.085
2020:03     -54.385   0.499   1.426   1.085   0.769
2020:04      17.704 -53.615   0.514   1.309   1.046
2021:01       3.908  13.681 -45.340   0.535   1.204
2021:02       1.536   2.962  10.570 -34.528   0.594
2021:03       2.737   1.174   2.251   8.165 -24.052
2021:04       5.255   1.998   0.926   1.717   6.305
2022:01       3.294   3.328   0.586   0.757   1.315
2022:02       3.272   1.938   1.134   0.019   0.641
2022:03       2.126   1.992   1.119  -0.283  -0.263
2022:04       2.255   1.312   1.211   0.737  -0.927
2023:01       1.635   1.549   0.898   0.831   0.608
2023:02       1.671   0.893   1.118   0.729   0.685
2023:03       1.274   1.136   0.484   0.901   0.681
2023:04       1.147   0.862   0.731   0.348   0.810
2024:01       1.348   0.942   0.676   0.566   0.400
2024:02       1.731   1.165   0.822   0.642   0.552
2024:03          NA   1.552   1.044   0.811   0.703
2024:04          NA      NA   1.332   1.005   0.849
2025:01          NA      NA      NA   1.214   1.016
2025:02          NA      NA      NA      NA   1.163
```

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: EMP (Nonfarm Payroll Employment)
By Forecast Step (1 to 5)
Transformation: Q/Q Growth Rate

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

Last Updated: 05/24/2024 12:50

_____

Qtr Forecast Step 1  Step 2  Step 3  Step 4  Step 5
2017:03       1.356   1.538   1.419   1.738   1.285
2017:04       1.295   1.356   1.538   1.419   1.738
2018:01       1.510   1.295   1.356   1.538   1.419
2018:02       1.708   1.510   1.295   1.356   1.538
2018:03       1.761   1.708   1.510   1.295   1.356
2018:04       1.686   1.761   1.708   1.510   1.295
2019:01       1.745   1.686   1.761   1.708   1.510
2019:02       1.687   1.745   1.686   1.761   1.708
2019:03       1.179   1.687   1.745   1.686   1.761
2019:04       1.360   1.179   1.687   1.745   1.686
2020:01       1.639   1.360   1.179   1.687   1.745
2020:02       0.758   1.639   1.360   1.179   1.687
2020:03     -39.985   0.758   1.639   1.360   1.179
2020:04      22.908 -39.985   0.758   1.639   1.360
2021:01       5.167  22.908 -39.985   0.758   1.639
2021:02       2.066   5.167  22.908 -39.985   0.758
2021:03       4.808   2.066   5.167  22.908 -39.985
2021:04       6.677   4.808   2.066   5.167  22.908
2022:01       4.805   6.677   4.808   2.066   5.167
2022:02       4.739   4.805   6.677   4.808   2.066
2022:03       3.398   4.739   4.805   6.677   4.808
2022:04       3.098   3.398   4.739   4.805   6.677
2023:01       2.514   3.098   3.398   4.739   4.805
2023:02       2.482   2.514   3.098   3.398   4.739
2023:03       1.798   2.482   2.514   3.098   3.398
2023:04       1.571   1.798   2.482   2.514   3.098
2024:01       1.624   1.571   1.798   2.482   2.514
2024:02       1.982   1.624   1.571   1.798   2.482
2024:03          NA   1.982   1.624   1.571   1.798
2024:04          NA      NA   1.982   1.624   1.571
2025:01          NA      NA      NA   1.982   1.624
2025:02          NA      NA      NA      NA   1.982
```

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: EMP (Nonfarm Payroll Employment)
By Forecast Step (1 to 5)
Transformation: Q/Q Growth Rate
Lag Length for DAR(p): AIC

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

Last Updated: 05/24/2024 12:50

Qtr Forecast	Step 1	Step 2	Step 3	Step 4	Step 5
2017:03	1.263	1.182	0.896	1.203	0.748
2017:04	1.168	1.136	1.250	0.780	0.925
2018:01	1.453	1.064	1.036	1.116	0.918
2018:02	1.605	1.340	0.983	0.964	0.997
2018:03	1.686	1.504	1.193	0.985	0.976
2018:04	1.556	1.563	1.403	1.087	0.973
2019:01	1.578	1.425	1.387	1.246	1.050
2019:02	1.539	1.423	1.333	1.237	1.114
2019:03	1.045	1.389	1.354	1.197	1.148
2019:04	1.205	0.885	1.314	1.223	1.149
2020:01	1.622	1.125	0.861	1.196	1.176
2020:02	0.676	1.567	1.236	0.815	1.167
2020:03	-54.385	0.531	1.360	1.167	1.020
2020:04	17.704	-27.453	0.453	1.243	1.063
2021:01	3.908	47.861	-21.273	0.473	1.110
2021:02	1.536	2.957	11.503	-15.104	0.869
2021:03	2.737	0.210	2.614	8.477	-8.891
2021:04	5.255	2.695	0.704	1.871	5.268
2022:01	3.294	2.966	0.898	0.791	1.298
2022:02	3.272	1.650	0.682	0.549	0.640
2022:03	2.126	1.704	0.805	-0.048	1.207
2022:04	2.255	0.970	0.975	0.210	0.214
2023:01	1.635	1.365	0.962	0.376	0.001
2023:02	1.671	0.986	1.013	0.562	-0.120
2023:03	1.274	1.363	1.027	0.680	0.215
2023:04	1.147	0.905	1.069	0.735	0.375
2024:01	1.348	1.100	1.020	0.802	0.512
2024:02	1.731	1.214	1.043	0.869	0.608
2024:03	NA	1.611	1.052	0.929	0.760
2024:04	NA	NA	1.307	1.018	0.850
2025:01	NA	NA	NA	1.147	0.945
2025:02	NA	NA	NA	NA	1.056

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: EMP (Nonfarm Payroll Employment)
By Forecast Step (1 to 5)
Transformation: Q/Q Growth Rate
Lag Length for DARM(p):  AIC

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

Last Updated: 05/24/2024 12:50

_____
```

Qtr Forecast	Step 1	Step 2	Step 3	Step 4	Step 5
2017:03	1.656	1.515	1.164	0.748	0.688
2017:04	1.373	1.304	1.394	1.120	0.696
2018:01	1.502	1.595	1.204	1.602	1.195
2018:02	1.439	1.448	1.615	0.916	1.471
2018:03	1.629	1.607	1.263	1.997	0.862
2018:04	1.717	1.341	1.219	1.257	1.906
2019:01	2.039	1.942	1.024	1.395	1.187
2019:02	1.546	1.858	1.646	0.900	1.210
2019:03	1.185	1.399	1.788	1.895	0.763
2019:04	1.251	0.850	1.598	1.639	1.655
2020:01	1.607	1.236	0.817	1.194	1.526
2020:02	-43.586	1.670	0.949	0.555	1.342
2020:03	31.374	-41.703	1.524	0.980	0.798
2020:04	22.811	26.441	-54.355	1.556	0.806
2021:01	19.407	6.721	8.867	-49.493	1.392
2021:02	-8.908	2.056	2.594	27.393	-44.710
2021:03	5.399	1.181	1.002	0.950	35.400
2021:04	13.617	5.287	1.977	-0.551	-2.775
2022:01	4.126	3.738	3.164	1.300	-0.169
2022:02	3.415	2.996	1.636	1.811	1.467
2022:03	3.422	2.988	2.110	0.595	0.896
2022:04	2.075	2.221	2.215	1.983	0.612
2023:01	3.087	1.516	1.293	2.197	1.658
2023:02	1.667	1.760	1.087	1.336	2.334
2023:03	1.389	1.354	0.993	0.524	0.878
2023:04	1.213	0.944	0.802	0.600	0.211
2024:01	1.993	1.028	0.741	0.523	1.109
2024:02	1.788	1.597	0.821	0.751	0.761
2024:03	NA	1.540	1.296	0.251	1.073
2024:04	NA	NA	1.336	0.636	0.012
2025:01	NA	NA	NA	0.903	0.505
2025:02	NA	NA	NA	NA	0.671

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: EMP (Nonfarm Payroll Employment)
Transformation: Q/Q Growth Rate

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

Last Updated: 05/24/2024 12:50

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)
2017:03	1.295	1.520	1.457	1.483	1.399
2017:04	1.510	1.524	1.476	1.462	1.244
2018:01	1.708	1.727	1.751	1.671	1.785
2018:02	1.761	1.725	1.890	1.763	1.949
2018:03	1.686	1.805	1.526	1.541	1.360
2018:04	1.745	1.750	1.308	1.346	1.160
2019:01	1.687	1.655	1.301	1.209	1.318
2019:02	1.179	1.166	1.139	1.109	1.485
2019:03	1.360	1.471	1.427	0.936	1.210
2019:04	1.639	1.672	1.643	1.303	1.302
2020:01	0.758	0.352	0.348	0.814	0.296
2020:02	-39.985	-39.991	-39.989	-39.811	-39.734
2020:03	22.908	23.333	21.549	22.086	22.317
2020:04	5.167	5.089	5.813	5.888	5.751
2021:01	2.066	2.080	3.592	2.973	2.702
2021:02	4.808	4.831	4.105	4.395	4.559
2021:03	6.677	4.754	5.681	5.851	5.851
2021:04	4.805	4.878	5.469	5.388	5.388
2022:01	4.739	4.711	4.606	NA	4.299
2022:02	3.398	3.303	3.213	NA	3.275
2022:03	3.098	3.430	3.499	NA	3.499
2022:04	2.514	2.495	2.182	NA	2.182
2023:01	2.482	2.527	NA	NA	2.353
2023:02	1.798	1.729	NA	NA	1.957
2023:03	1.571	1.724	NA	NA	1.724
2023:04	1.624	1.587	NA	NA	1.587
2024:01	1.982	NA	NA	NA	1.982
2024:02	NA	NA	NA	NA	NA
2024:03	NA	NA	NA	NA	NA
2024:04	NA	NA	NA	NA	NA
2025:01	NA	NA	NA	NA	NA
2025:02	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: 2004:04-2021:04

SPF Projections for Nonfarm Payroll Employment, Transformation: Q/Q Growth Rate
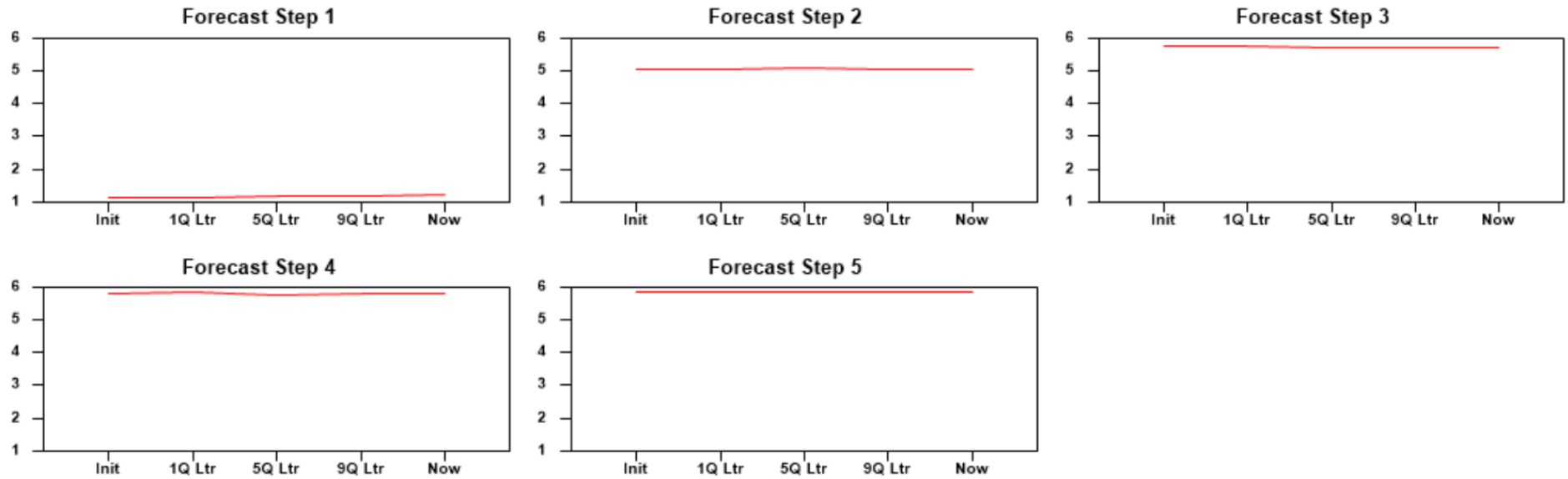
The RMSE (y-axis) is plotted for alternative measures of the realization (x-axis). T.Stark, FRB Phila.
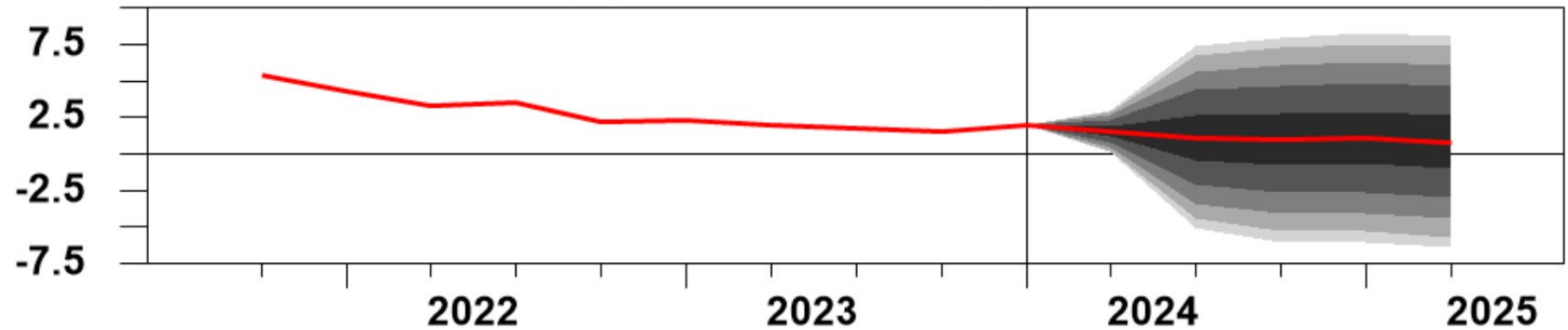
Nonfarm Payroll Employment

History, Forecasts, and Ranges for the SPF of 2024:02

Q/Q Growth Rate

(Ranges Cover 25 to 80 Percent Confidence)

Ranges use the N(0,MSE) density. MSE sample: 03:04-22:04 / Realization: Five Qtrs After Initial Release. T.Stark, FRB Phila.