Federal Reserve Bank of Philadelphia

Research Brief
Research Department

January 2021
https://doi.org/10.21799/frbp.rb.2021.jan

Battle of the Forecasts: Growth of the Median or Median of Growth as the SPF Consensus

Patrick Doelp
Federal Reserve Bank of Philadelphia Research Department

Fatima Mboup
Federal Reserve Bank of Philadelphia Research Department

Battle of the Forecasts: Growth of the Median or Median of Growth as the SPF Consensus

by **Patrick Doelp** and **Fatima Mboup**

In the nearly 30 years that the Federal Reserve Bank of Philadelphia has been conducting the widely followed Survey of Professional Forecasters (SPF), our analysis of the results for each new quarterly survey has emphasized the growth rates of the median forecast for levels.[1] The forecast growth rates reported in our SPF results are computed by first finding the median of the panelists' projections for the level of a variable. We then compute a consensus growth forecast as the growth of the consensus level, a procedure we call "growth of the median," or GM. Alternatively, we could first compute each panelist's growth forecast and report a consensus projection as the median of the growth rates, which we call "the median of the growth rates" method, or MG. We have always assumed, without formal analysis, that the GM method produces a better consensus forecast than the MG method.

In this Research Brief we formally study which method, MG or GM, produces a more accurate consensus projection for growth. We focus our analysis on nine variables included in the survey. We want our results to be as robust as possible, so we examine five forecast horizons and four alternative measures of the realizations from which we compute the forecast errors, and we check five alternative sample periods. We apply the well-known Diebold–Mariano (1995) statistical test for relative forecast accuracy between the growth rate of the median projections (GM) and the projections for median of the growth rates (MG).

Overall, our analysis shows very few statistically significant differences between the two methods for computing consensus growth-rate projections. Even when we find a statistically significant difference between the methods, the measured difference in forecast accuracy is so small that we regard the result as not very meaningful. Our main conclusion is that either method for computing a consensus forecast for growth works reasonably fine.

Patrick Doelp and **Fatima Mboup** are research associates at the Federal Reserve Bank of Philadelphia.

The views expressed by the authors are not necessarily those of the Federal Reserve.

Patrick T. Harker
PRESIDENT AND
CHIEF EXECUTIVE OFFICER

Michael Dotsey
EXECUTIVE VICE PRESIDENT
AND DIRECTOR OF RESEARCH

Adam Steinberg
MANAGING EDITOR,
RESEARCH PUBLICATIONS

Brendan Barry
DATA VISUALISATION
MANAGER

1 For more information on the Survey of Professional Forecasters, see www.philadelphiafed.org/research-and-data/real-time-center/survey-of-professional-forecasters.

Real-Time Methodological Considerations

Currently 23 variables are forecast in the SPF. In this paper, we focus on industrial production, nominal GDP, the GDP price index, real total consumption expenditures, real federal government consumption and investment, real GDP, real nonresidential fixed investment, real residential fixed investment, and real state and local consumption and investment. In each quarterly survey, the panelists provide their projections for the values the variables will take in the quarter when we conduct the survey and in the following four quarters. These five alternative horizons provide five alternative ways in which we can test our hypothesis about forecast accuracy.

Our study pays particular attention to the real-time nature of our statistical experiments. Recognizing that the SPF is conducted in real time, with the panelists using only information available at the time, we must exercise some caution in our choice of historical realizations when assessing forecast accuracy. Given that the U.S. statistical agencies often revise the historical realizations, we focus our comparisons of the accuracy of GM and MG forecasts to the first, second, third, and most recent releases of each variable. These alternative ways to measure realizations add to the number of ways in which we can test our hypothesis.

Because the SPF began in 1968, our analysis focuses on the period from 1968 Q4 to 2018 Q1. We analyze our results over the full sample period, 1968 to 2018, and also over the following sample periods to ensure the results are robust: 1968 Q4 to 1990 Q1, 1990 Q1 to 2018 Q1, 2000 Q1 to 2018 Q1, 2005 Q1 to 2018 Q1, and 2010 Q1 to 2018 Q1. The year 1991 marks the beginning of the Philadelphia Fed's tenure for conducting the SPF. Prior to 1990, the National Bureau of Economic Research (NBER) and the American Statistical Association (ASA) conducted the survey. The difference in survey ownership prompted our interest in analyzing our results over sample periods after the Philadelphia Fed took over the survey. The sample period from 2005 to 2018 covers the Great Recession, and the period from 2010 to 2018 covers the subsequent recovery.

Defining Growth Rates and Testing Forecast Accuracy

We define a "growth in median" consensus forecast (GM) as the growth in the median consensus forecast for the level of a variable (X) from one quarter to the next, according to

$$\widehat{GM}_{t+k|t-1} = 100\left[\left(\frac{\hat{X}_{t+k|t-1}}{\hat{X}_{t+k-1|t-1}}\right)^4 - 1\right], \quad k = 0,1,\ldots,4$$

where $\widehat{GM}_{t+k|t-1}$ represents the forecast (made on the basis of observations known through period $t-1$) for quarter-over-quarter growth in period $t+k$; k represents the forecasted horizon in quarters; and $\hat{X}_{t+k|t-1}$ represents the corresponding median forecast for the level.

Our computation of the "median of growth" consensus projection (MG) uses the median of the panelists' growth projections, according to

$$\widehat{MG}_{t+k\,|\,t-1} = med\left\{\hat{g}^{1}_{t+k|t-1},\ \hat{g}^{2}_{t+k|t-1},\ \hat{g}^{3}_{t+k|t-1}\ldots,\ \hat{g}^{N}_{t+k|t-1}\right\},\quad k=0,1,\ldots,4$$

$$where\ \hat{g}^{i}_{t+k|t-1} = 100\left[\left(\frac{\hat{x}^{i}_{t+k|t-1}}{\hat{x}^{i}_{t+k-1|t-1}}\right)^{4}-1\right],\ i=1,2,\ldots N$$

where i indexes an individual panelist and N denotes the number of panelists (which might differ from one survey to the next); $\hat{g}^{i}_{t+k\,|\,t-1}$ represents panelist i's growth-rate projection; $\hat{x}^{i}_{t+k\,|\,t-1}$ represents panelist i's prediction for the level of the variable; and $med\left\{\ \right\}$ denotes the median.

When testing for relative accuracy of the competing SPF forecast growth rates, we follow the methodology of Stark (2010) in using the conventional mean-square-error (MSE) statistic as the measure of accuracy. We define the mean of the squared forecast errors according to

$$mse_{j}\left(\tau,r\right) = \frac{1}{T}\sum_{t}\left(\varepsilon^{(r)}_{j,\ t+\tau|t}\right)^{2},\ \ \tau=0,\ldots,4;\ \ r=1,\ldots,4;\ j=1,2$$

where $\varepsilon^{(r)}_{j,\ t+\tau|t}$ is the τ-quarter-ahead forecast error, defined on the r^{th} measure of the realization; j is an index to distinguish between GM and MG forecasts; and T measures the number of forecast errors in the sample period.

We use the Diebold–Mariano (1995) statistic to test whether the GM forecasts are just as accurate as the MG forecasts. Formally, we follow Diebold and Mariano (1995) in designing the null and alternative hypotheses as

$$H_{0}: mse_{MG}\left(\tau,r\right) = mse_{GM}\left(\tau,r\right)$$

$$H_{1}: mse_{MG}\left(\tau,r\right) \neq mse_{GM}\left(\tau,r\right)$$

where, as noted above, the mse_{MG} refers to the MSE for MG while mse_{GM} refers to the MSE for GM. We conduct a separate test for each variable, forecast horizon $(\tau=0,1,\ldots,4)$, realization $(r=1,\ldots,4)$, and sample period.

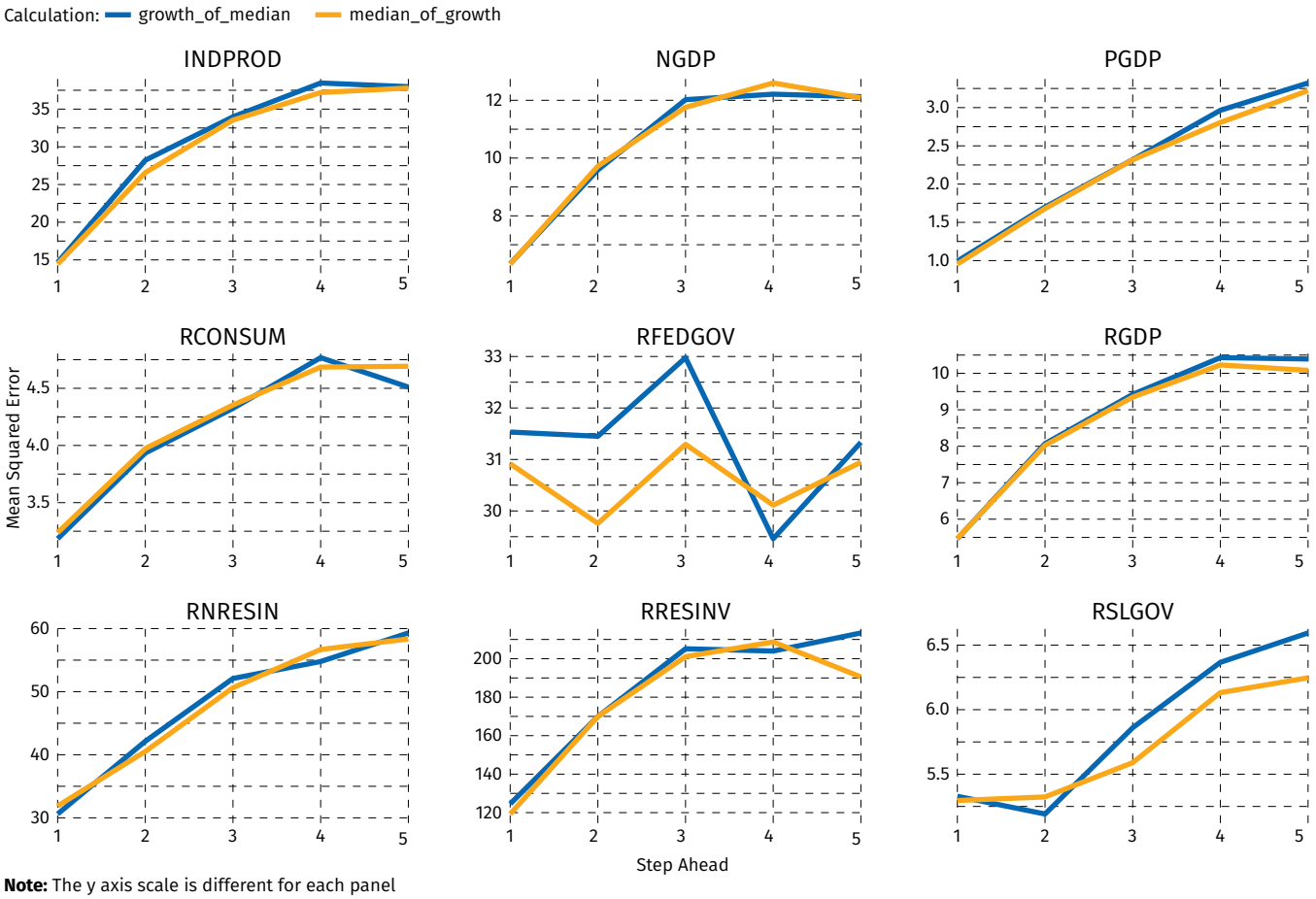
Statistical Findings

We begin with a graphical analysis of the differences in forecast accuracy. Figure 1 presents graphs of the mean square error (MSE) for the competing "growth in median" (GM) and "median of growth" (MG) projections. We focus on the full sample period using the most recent historical values for realizations. Each graph shows the MSE (y-axis) against the five forecast horizons (x-axis). Although differences exist between the GM and MG forecasts' accuracy, the differences are small except for those associated with federal government spending and with real state and local government spending. We found similar results using alternative measures of realizations and sample periods.[2]

FIGURE 1

MSE Results

Release: most_recent (full sample)



Note: The y axis scale is different for each panel

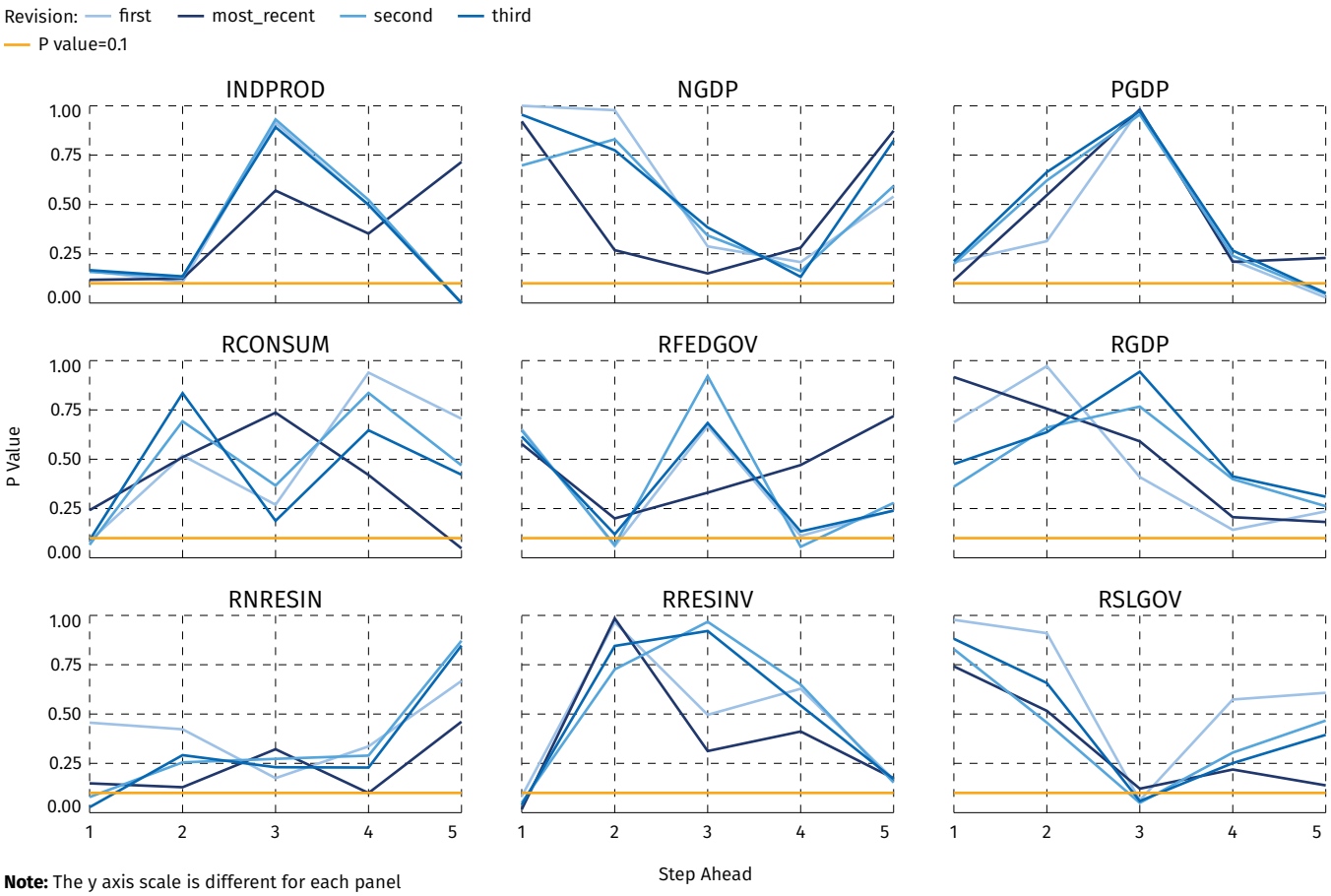
2 Results for different sample periods and alternative realizations are available upon request.

To analyze the accuracy of the GM and MG forecasts more rigorously and formally, we use the Diebold–Mariano (DM) test.[3] Figure 2 presents the p-values for the DM test in graphical form, using a red cutoff line for a 10 percent significance level. (Values below the 10 percent cutoff indicate a statistically significant result.) The other four colored lines represent the use of alternative realizations for constructing the p-values for the hypothesis tests (y-axis) at each quarterly forecast horizon (x-axis). Figure 2 suggests that, on average across all variables and forecast horizons, there is no statistically significant difference between GM and MG forecast accuracy. In the few cases where a statistical significance exists, the DM test statistic is small in comparison to the MSE for GM and MG, indicating that the presence of statistical significance does not imply economic relevance. We found similar results when analyzing across alternative sample periods.

FIGURE 2

Diebold—Mariano Test Results

1968 Q4–2018 Q1



Note: The y axis scale is different for each panel

3 Our Diebold–Mariano statistics reflect the usual heteroscedasticity and autocorrelation consistent (HAC) correction using a flat window and a Newey–West window. The latter correction is used when the variance is less than or equal to zero. We also use a truncation lag parameter of k – 1 where k = 0, and 1...4 indicates the forecast steps. We employ a 10 percent significance level.

Table 1 provides additional details for some illustrative, but informative, cases. All of the statistics presented in Table 1 are for cases where the DM test statistic indicated statistical significance. Looking closely at the results for real federal government consumption and investment (RFEDGOV) over the entire sample period, we see that the DM test statistic is sizable, but so are the MSEs for the GM and MG projections. In other words, both forecasts, GM and MG, are inaccurate.

TABLE 1

Variable	Revision	Forecast Horizon	P-Value	DM Test	Sample Period	MSE_GM	MSE_MG
INDPROD	first	5	0.000228003	1.063210223	1968 Q4–2018 Q1	41.62077413	42.68398435
INDPROD	second	5	6.47E-06	0.992176456	1968 Q4–2018 Q1	40.86004754	41.852224
INDPROD	third	5	0.00018807	0.956059756	1968 Q4–2018 Q1	40.99363353	41.94969329
PGDP	first	5	0.02805524	-0.12753504	1968 Q4–2018 Q1	3.501499049	3.373964008
PGDP	second	5	0.045382154	-0.128478039	1968 Q4–2018 Q1	3.938519305	3.810041266
PGDP	third	5	0.050304059	-0.122417215	1968 Q4–2018 Q1	3.852858877	3.730441663
RCONSUM	first	1	0.085268804	0.096103298	1968 Q4–2018 Q1	3.155507998	3.251611296
RCONSUM	second	1	0.066122224	0.100742343	1968 Q4–2018 Q1	3.029455535	3.130197878
RCONSUM	third	1	0.090200541	0.101670346	1968 Q4–2018 Q1	3.134553426	3.236223772
RCONSUM	most_recent	5	0.048037105	0.181752905	1968 Q4–2018 Q1	4.51087273	4.692625635
RFEDGOV	first	2	0.059584397	6.333562992	1968 Q4–2018 Q1	85.24868728	91.58225028
RFEDGOV	second	2	0.065263822	5.844027285	1968 Q4–2018 Q1	92.71065661	98.5546839
RFEDGOV	second	4	0.056029207	2.818217436	1968 Q4–2018 Q1	95.9714187	98.78963613
RNRESIN	second	1	0.079009043	0.767354078	1968 Q4–2018 Q1	48.34113565	49.10848973
RNRESIN	third	1	0.027539869	1.068192203	1968 Q4–2018 Q1	48.7415618	49.80975401
RNRESIN	most_recent	4	0.099537749	1.901821224	1968 Q4–2018 Q1	54.78093907	56.68276029
RRESINV	first	1	0.082647382	-2.439355882	1968 Q4–2018 Q1	113.8475116	111.4081557
RRESINV	second	1	0.05069218	-3.299541634	1968 Q4–2018 Q1	111.269346	107.9698044
RRESINV	third	1	0.039457881	-5.051581048	1968 Q4–2018 Q1	123.1284722	118.0768912
RRESINV	most_recent	1	0.016652323	-5.262527051	1968 Q4–2018 Q1	124.5355386	119.2730116
RSLGOV	first	3	0.062475907	-0.307807828	1968 Q4–2018 Q1	4.683791613	4.375983785
RSLGOV	second	3	0.049566364	-0.329826047	1968 Q4–2018 Q1	5.680891779	5.351065731
RSLGOV	third	3	0.058725567	-0.335199901	1968 Q4–2018 Q1	5.77721162	5.442011719

Conclusion

For over 30 years we have emphasized the growth rate of the median forecasts in our SPF results. In this research brief we formally test the accuracy of the growth rate of the median forecasts (GM) against an alternative measure for forecasted growth rates, the median of forecasted growth rates (MG). We have always used the former method in order to safeguard against possible typographical errors and outliers in the data set. Our formal analysis concludes that generally the difference in accuracy between the GM and MG forecasts in the SPF is statistically minor. A few cases suggested statistically major differences in forecast accuracy. However, we also find that these statistically major cases are of little economic relevance. We therefore conclude that the survey's forecasted growth rate of the median forecast (GM) is just as accurate as the median of the forecasted growth rate (MG). RB

References

Diebold, Francis X., and Robert S. Mariano. "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics*, 13:3 (July 1995), pp. 253–263.

Stark, Tom. "Realistic Evaluation of Real-Time Forecasts in the Survey of Professional Forecasters," Federal Reserve Bank of Philadelphia *Research Rap Special Report* (May 28, 2010).