

**Finance and Economics Discussion Series  
Divisions of Research & Statistics and Monetary Affairs  
Federal Reserve Board, Washington, D.C.**

**Raiders of the Lost High-Frequency Forecasts: New Data and  
Evidence on the Efficiency of the Fed's Forecasting**

**Andrew C. Chang and Trace J. Levinson**

**2020-090**

Please cite this paper as:

Chang, Andrew C., and Trace J. Levinson (2020). "Raiders of the Lost High-Frequency Forecasts: New Data and Evidence on the Efficiency of the Fed's Forecasting," Finance and Economics Discussion Series 2020-090. Washington: Board of Governors of the Federal Reserve System, <https://doi.org/10.17016/FEDS.2020.090>.

NOTE: Staff working papers in the Finance and Economics Discussion Series (FEDS) are preliminary materials circulated to stimulate discussion and critical comment. The analysis and conclusions set forth are those of the authors and do not indicate concurrence by other members of the research staff or the Board of Governors. References in publications to the Finance and Economics Discussion Series (other than acknowledgement) should be cleared with the author(s) to protect the tentative character of these papers.

# Raiders of the Lost High-Frequency Forecasts: New Data and Evidence on the Efficiency of the Fed’s Forecasting

Andrew C. Chang\*      Trace J. Levinson†

October 23, 2020

## Abstract

We introduce a new dataset of real gross domestic product (GDP) growth and core personal consumption expenditures (PCE) inflation forecasts produced by the staff of the Board of Governors of the Federal Reserve System. In contrast to the eight Greenbook forecasts a year the staff produces for Federal Open Market Committee (FOMC) meetings, our dataset has roughly weekly forecasts. We use these new data to study whether the staff forecasts efficiently and whether efficiency, or lack thereof, is time-varying. Prespecified regressions of forecast errors on forecast revisions show that the staff’s GDP forecast errors correlate with its GDP forecast revisions, particularly for forecasts made more than two weeks from the start of a FOMC meeting, implying GDP forecasts exhibit time-varying inefficiency between FOMC meetings. We find some weaker evidence for inefficient inflation forecasts.

- *Keywords:* Federal Reserve; Forecast efficiency; Information Rigidities; High-frequency forecasts; Preanalysis plan; Preregistration plan; Real-time data.
- *JEL Codes:* C53; C82; D79; E27; E37; E58.

---

\*Corresponding author. Principal Economist, Division of Research and Statistics, Board of Governors of the Federal Reserve System. a.christopher.chang@gmail.com. <https://sites.google.com/site/andrewchristopherchang/>. ORCID: 0000-0002-9769-789X.

†Former Senior Research Assistant, Division of Research and Statistics, Board of Governors of the Federal Reserve System.

‡You can download code and data for this paper at [https://www.federalreserve.gov/econres/feds/files/Chang-and-Levinson\\_high-frequency-forecasts\\_Replication.zip](https://www.federalreserve.gov/econres/feds/files/Chang-and-Levinson_high-frequency-forecasts_Replication.zip). The views and opinions expressed here are ours and are not necessarily those of the Board of Governors of the Federal Reserve System. This paper makes use of forecasts by the Board’s staff other than those from the Greenbook/Tealbook. These non-Greenbook/non-Tealbook forecasts provided updates to the staff forecast in-between the Greenbooks/Tealbooks that incorporated information from data released subsequent to the most recent Greenbook/Tealbook. However, the staff typically, but not always, updated conditioning assumptions regarding fiscal policy, monetary policy, financial conditions, or other economic forces that shape the trajectory of the economy only for the Greenbook/Tealbook. The forecasts that played the most direct role in monetary policymaking were the forecasts from the Greenbook/Tealbook.

# 1 Introduction

The staff of the Board of Governors of the Federal Reserve System (the staff) prepare a detailed set of forecasts, both of the U.S. and international economies, for each regularly scheduled Federal Open Market Committee (FOMC) meeting. These forecasts are called “Greenbook” forecasts.<sup>1</sup> However, Greenbook forecasts are only some of the analysis of the economy by the staff for the Board of Governors. Since there is new information about the economy every day, the staff provides new forecasts to the Board of Governors between Greenbooks. However, these between-Greenbook forecasts have not previously been usable for study by researchers.

We introduce these between-Greenbook forecasts and use a preanalysis plan to study their statistical efficiency. Our paper expands on research that improves the availability of real-time forecast data, such as Croushore & Van Norden (2018), and also expands on literature that evaluates the real-time performance of professional forecasters, such as Arai (2016).

We assemble a unique set of Federal Reserve documents to construct a new dataset that contains these between-Greenbook forecasts, which we call our “high-frequency forecast dataset.” Roughly speaking, we have weekly forecasts of annualized quarterly real gross domestic product (GDP) growth and annualized quarterly core personal consumption expenditures (PCE) inflation from 2001 to 2011 for one-quarter backcasts through two-quarter ahead forecasts.

Although the staff monitors economic news every day, it does not have infinite forecasting capacity. Forecasts that are made just before a FOMC meeting receive more attention from the Board of Governors —particularly Greenbook forecasts —as they play the most direct role in monetary policy. The staff often uses time just after a FOMC meeting for projects that enhance its productivity and long-run relevance to the mandates of the Federal Reserve,

---

<sup>1</sup>Since 2010 these forecasts have been called “Tealbooks”. For simplicity we refer to Tealbooks as Greenbooks.

such as the following: developing models, economic research, investigating the usefulness of new data for studying the economy, or upgrading infrastructure.

A natural hypothesis is that the staff allocates more resources to the forecasts that are made just before a FOMC meeting and, therefore, that these forecasts are better. Due to the high-frequency nature of our new data, we are able to test this hypothesis. Using prespecified regressions of forecast errors on forecast revisions, we find that GDP forecasts made in the two weeks before a FOMC meeting are efficient —forecast revisions do not predict forecast errors. But GDP forecasts made more than two weeks from a meeting are inefficient —the staff’s current-quarter GDP forecast revisions are positively correlated with its forecast errors, which suggests that the staff *underrevises* these current-quarter GDP forecasts. This underrevision is consistent with the staff following an anchoring heuristic (Tversky & Kahneman (1974), Kahneman & Tversky (1977), Campbell & Sharpe (2009)) where the staff adheres too closely to its previous current-quarter GDP forecast after receiving new macroeconomic information. This underrevision could be due to information rigidities, which Coibion & Gorodnichenko (2012) document for professional forecasters, that take significant effort to overcome. That said, the staff’s two-quarter ahead forecast revisions are negatively correlated with its forecast errors, which suggest that the staff *overrevises* these two-quarter ahead GDP forecasts. For inflation, we find some evidence that the staff forecasts also *overrevise*, but this evidence is less strong.

While forecast efficiency regressions can speak to the sub-optimality forecasts, they are silent on what information could be processed better to make optimal forecasts. We examine one fountain of information aggregation —financial markets —and investigate whether the staff could use the reaction of financial markets to macroeconomic news to improve its forecasts.

We prespecify one measure of how financial markets react to news: changes in S&P 500 futures returns following the release of macroeconomic news (for example, the employment

report) weighted by the size of the news.<sup>2</sup> Increases in S&P 500 futures returns following a news release suggest that the news was better for the economy than the market expected. We find that our return-weighted macroeconomic news measure predicts the staff's GDP forecast errors, implying that the staff may not efficiently account for information in asset price changes to assess macroeconomic news which, again, could be due to information rigidities. This evidence also suggests that if FOMC announcements are influenced by the staff forecasts, then market reactions to FOMC announcements are likely due to new information about monetary policy, a view supported by Bernanke & Kuttner (2005), Woodford (2005), and Bauer & Swanson (2020), and are less likely to be due to new information about the economy, which is at odds with the inferences of Romer & Romer (2000), Campbell, Evans, Fisher & Justiniano (2012), and Nakamura & Steinsson (2018).

## 2 A New High-Frequency Forecast Dataset

Greenbooks are comprehensive forecasts of the U.S. and global economies. Greenbooks contain forecasts of: several U.S. macroeconomic indicators (for example: GDP, payroll gains, unemployment rate, and inflation), U.S. financial market indicators (for example: the 10-year treasury yield, equity prices, and corporate profits), non-U.S. aggregates, the output gap, and the non-accelerating inflation rate of unemployment (NAIRU), among others. Typically, Greenbook forecasts cover a horizon of at least five quarters, though the exact horizon has varied over time.

The staff produce the Greenbooks for each regularly scheduled FOMC meeting, which, in recent years, occur eight times per year. Given their importance to monetary policy, the literature has closely evaluated the Greenbook forecasts.<sup>3</sup>

---

<sup>2</sup>We measure macroeconomic news as the standardized difference between a macroeconomic data release and its median eve-of-release forecast from the panel of economists surveyed by Bloomberg (Bloomberg Finance LP 2017). This type of macroeconomic news measurement is common in the literature. For example, Scotti (2016) defines macroeconomic news similarly.

<sup>3</sup>Examples include the following: Joutz & Stekler (2000), Romer & Romer (2008), Tulip (2009), Ericsson, Hood, Joutz, Sinclair & Stekler (2015), Messina, Sinclair & Stekler (2015), Chang & Hanson (2016),

While Greenbook forecasts are relatively accurate, they are available only immediately before the eight regularly scheduled FOMC meetings per year.<sup>4,5</sup> But the macroeconomy evolves continuously and new macroeconomic data are available daily. Therefore, the staff communicates its understanding of economy and its interpretation of new macroeconomic data to the Board of Governors much more frequently than eight times per year. The exact timing and nature of these communications varies over time, reflecting both state of the economy and the preferences of Board members. In recent years, typical forms of communication to the Board of Governors outside of the Greenbooks include written forecast update memos (approximately weekly) and regularly scheduled (approximately semi-monthly) in-person forecast update briefings. These non-Greenbook forums allow the staff to provide the Board of Governors high-frequency updates on the staff’s view of the economy. These updates are usually short-horizon forecasts that do not cover the same range of macroeconomic data nor are vetted by the staff to the same degree as the Greenbooks.

We collected data from archived documents of the staff’s communications to the Board of Governors, such as the forecast update memos and the in-person forecast updates, to construct our high-frequency forecast dataset. The six types of archived documents that we used were:

1. Briefing texts. When the staff conducts an in-person forecast update briefing for the Board of Governors, a staff member first delivers a set of prepared remarks. The Board of Governors follow this delivery with questions. Some of the texts contain forecasts.
2. Briefing tables and charts. For each in-person forecast update briefing, the staff creates a set of tables and charts to accompany the prepared remarks. Some of these tables and charts contain the staff forecasts.

---

Croushore & Van Norden (2018), Berge, Chang & Sinha (2019), Croushore & Van Norden (2019), Reifschneider & Tulip (2019).

<sup>4</sup>On Greenbook accuracy, see Joutz & Stekler (2000), Romer & Romer (2008), or Chang & Hanson (2016).

<sup>5</sup>In recent years, regularly scheduled FOMC meetings occur in the first and third month of each quarter, although the staff produce forecasts for irregular meetings of the FOMC. Irregular meetings occur as a reaction to extreme events.

3. Eve-of-GDP-release database snapshots. The staff sometimes saves a snapshot of its forecasts for real GDP growth and core PCE inflation on the eve-of-release of the official Bureau of Economic Analysis (BEA) estimate for GDP. These snapshots are called “killsheets” or “comparison sheets.”
4. Irregular database backups. We use archived automatic backups of databases that contain the staff’s GDP and inflation forecasts, called “RUTH backups” or “weekly RUTHs.” These database backups were irregular as they: (1) occurred at different frequencies over time, (2) were not necessarily at times that the staff vetted its forecasts, and (3) were not necessarily at times that the staff updated the Board of Governors.
5. Forecast update memos. In recent years the staff has delivered an approximately weekly written forecast update memo that summarized its thinking. These memos sometimes contained the numerical values of the staff’s GDP and inflation forecasts. The staff also delivers memos to the Board of Governors around extreme events or when asked for an update by a Board member.
6. Greenbooks.

We extracted forecasts of annualized quarterly real GDP growth and annualized quarterly core PCE inflation from these six types of archived documents. See appendix A for additional details.

Figure 1 introduces our dataset. We show the forecasts of annualized quarterly real GDP growth and core PCE inflation for: the previous calendar quarter (*backcasts*, top row), the current calendar quarter (*nowcasts*, middle row), and one-quarter ahead forecasts (bottom row). Because the staff’s forecasts are tied to the schedule of FOMC meetings, we indexed forecast horizons in our high-frequency forecast dataset based on when the forecasts are made as was relative at the next FOMC meeting.<sup>6</sup> In the left panels of Figure 1, each black

---

<sup>6</sup>This timing convention implied that the forecast horizon for forecasts made just after a FOMC but just before a calendar quarter changed could be lagged once. For example, if we observed a forecast of Q1

dot is a between-Greenbook forecast of GDP, which are roughly weekly. Green triangles are Greenbook forecasts; there are typically two of them per quarter. The right panels show the same information for inflation. The first forecast in our data is on January 22, 2001, and the final forecast is on December 23, 2011. National Bureau of Economic Research (NBER) recessions are shaded.

Figure 1 shows that the staff frequently updates its forecasts of real GDP growth and core PCE inflation between Greenbooks. For example, we have an average of 90 current-quarter forecasts of GDP per year, only eight of which per year are Greenbooks. We observed the most GDP forecasts in the mid-2000s, and somewhat fewer in the early and late years of our sample. Our data are more sparse for the staff's inflation projections, particularly for further ahead forecasts, as seen in the right panels of Figure 1. Early in the sample, we do not have inflation forecasts outside of Greenbooks. However, we see more between-Greenbook forecasts later in our sample. For example, in 2011 we have 77 current-quarter forecasts of inflation.

That the number of forecasts in our dataset differs across years and also differs between GDP and inflation is because both the staff's forecasting methods and recording standards evolved over time.<sup>7</sup> For changing recording standards, we gathered all archived documents on between-Greenbook forecasts that we are aware of. But because the staff was not intending for these documents to be used as a real-time account of its forecasts, it is possible that the staff was updating its between-Greenbook forecasts and did not record them in real time. This lack of intention may also be what limits our dataset to GDP and inflation. And, for database backups in particular, because the backups occurred at irregular intervals, the forecasts recorded by the backups may not have been fully vetted by the staff.

---

in March after the March FOMC when the next FOMC was in April, then we counted this forecast as a backcast.

<sup>7</sup>For example, starting with the February 17, 2000 Monetary Policy Report (Board of Governors of the Federal Reserve System 2000), the FOMC started characterizing inflation forecasts in terms of core PCE instead of the consumer price index (CPI). The shift in inflation metric, while before the beginning of our high-frequency forecast dataset, may have also been accompanied with a gradual shift in the frequency that the staff recorded its core PCE inflation forecasts.

Table 1 presents summary statistics of the forecast errors from our high-frequency forecast dataset, defined as the BEA’s third release minus the forecast, at each forecast horizon.<sup>8</sup> Figure 2 shows the current-quarter errors.

The staff’s forecasts are about unbiased, and the middle 70 percent of the current-quarter error distribution is also approximately symmetric around zero for both GDP and inflation. The exceptions to the unbiasedness are the two-quarter ahead forecasts. At two quarters ahead, the staff’s GDP projections were too high and the inflation projections were too low.

The staff’s forecasts are also quite variable and exhibit large misses across time. The range of current-quarter real GDP growth forecasts goes from a minimum of negative 7.1 percent (March 6, 2009) to a maximum of 5.5 percent (December 23, 2003). There is also notable variation within a particular quarter or year as macroeconomic data are realized. The forecasts of inflation also vary, albeit in a smaller range than GDP. Of course, our sample is relatively short: 11 years (44 quarters) and includes the Great Recession. That the staff made volatile forecasts and also made large errors in a recession is unsurprising, at least to us.

What is more surprising to us —and perhaps surprising to you as well —is that both the average GDP and inflation backcast and current-quarter errors were about zero during the Great Recession. The average GDP current-quarter error during the Great Recession was only 0.05 percentage points, whereas excluding the Great Recession it was 0.12 percentage points.<sup>9</sup> In contrast, the standard deviation of real GDP growth was more than 2.5 percentage points. Similarly, the average inflation current-quarter error during the Great Recession was only -0.12 percentage points whereas, excluding the Great Recession, it was less than one basis point.<sup>10</sup> The standard deviation of inflation was 0.64 percentage points.<sup>1112</sup> That

---

<sup>8</sup>The third release is available in the third month after the quarter ends, or about 25 weeks after the start of the quarter. See Landefeld, Seskin & Fraumeni (2008) or Chang & Li (2018) for a description of the BEA’s GDP revision process. Real-time data come from the real-time dataset for macroeconomists, documented by Croushore & Stark (2001) and Croushore & Stark (2003).

<sup>9</sup> $p = 0.78$  using a two-sided t-test with unequal variances.

<sup>10</sup> $p = 0.13$  using a two-sided t-test with unequal variances.

<sup>11</sup>Standard deviations calculated using BEA third-release estimates from 2001 to 2011.

<sup>12</sup>These results are at odds with the “missing disinflation” of the Great Recession where, conditional on

said, one-quarter ahead GDP forecasts were too high during the Great Recession, though inflation forecasts were still about right.

The novelty of our dataset is that it records the staff’s forecasts approximately weekly, whereas the staff makes Greenbook forecasts approximately every six weeks. However, our dataset comes with three important limitations. First, our dataset only has forecasts of real GDP growth and core PCE inflation, whereas the Greenbooks cover other macroeconomic and financial series. Second, our dataset covers one-quarter backcasts through two-quarter ahead forecasts and has a more limited sample further ahead, whereas Greenbooks consistently forecast five quarters ahead (sometimes more). Third, our dataset only extends back to 2001, whereas Greenbooks are available since 1967.

Nevertheless, our dataset can give us insight into how the staff incorporates real-time information into its GDP and inflation projections that Greenbooks cannot. Figure 3 shows the root mean squared error (RMSE) of the staff forecasts of GDP and inflation by week to the start of the quarter. Week zero denotes the start of a quarter (for example, the first week of January in Q1). Actual values are the BEA’s third estimates. The vertical dashed line represents the approximate release date of the BEA’s first estimate for real GDP growth and core PCE inflation, which occurs towards the end of the month after the quarter ends (for example, the end of April for a Q1 estimate).

Figure 3 shows that the staff continuously improves its GDP and inflation projections. And, while the staff’s RMSE steadily falls, the fall accelerates around the start of the quarter—evidence that either current-quarter data are particularly useful for the staff or that the staff devotes significant effort to continuously updating its nowcasts (or both). Furthermore, the staff’s RMSE continues to fall even after the BEA’s first estimates of GDP and inflation become available (approximately four weeks after the quarter ends), which indicates that the staff is nowcasting a BEA estimate of GDP and inflation that is released after the BEA’s

---

GDP during the Great Recession, the historical relationship between GDP and inflation would have predicted weaker inflation during the Great Recession than was the case. See Coibion & Gorodnichenko (2015) for a more detailed description of “missing disinflation”. The staff did not seem to think, given GDP, that disinflation was missing *in real time* once it started observing indicators of the current quarter.

first estimate.

### 3 Evaluating High-Frequency Forecasts

We now evaluate whether the staff’s forecasts satisfy standard efficiency tests.

For the results in this section, we used a preanalysis plan for our modeling, estimation, hypothesis testing, and data transformation choices. We used the plan to minimize our potential biases, to maximize the validity of our analysis, to avoid specification searching, and to avoid p-hacking. In particular, we thought that a preanalysis plan would particularly useful for mitigating our potential biases as, at the time of our initial writing in June 2017, we were involved in creating the Greenbook forecasts (although we were not involved in creating the forecasts in our dataset).<sup>13</sup>

We composed the initial draft and all revisions to our preanalysis plan after collecting the raw forecast data and making data cleaning choices but before finalizing our data cleaning programs and viewing the model estimation results in this section. The description in this section differs from the final version of our registered preanalysis plan only in terms of exposition —the models, estimators, data transformations, and so on are identical.<sup>14</sup>

#### 3.1 Average Efficiency —A Comparison with Previous Literature

Our first set of models provided a baseline comparison of our results of staff forecast bias and revision tendencies using our new high-frequency forecast dataset to existing research that only uses Greenbooks. These models were Mincer & Zarnowitz (1969) regressions of the staff’s forecast errors on its forecast revisions:

---

<sup>13</sup>Preanalysis plans in economics are uncommon. To our knowledge, Neumark (1999) and Neumark (2001) were the first economics studies that used preanalysis plans. Subsequent papers that used preanalysis plans include the following: Casey, Glennerster & Miguel (2012), Chang & Li (2017), Chang & Li (2018), Chang & Li (Forthcoming).

<sup>14</sup>We registered our initial preanalysis plan on March 13, 2017 with the Open Science Framework, and revised it three times on: April 19, 2017, May 5, 2017, and May 26, 2017. Our plan can be found at <https://osf.io/de3pe/>.

$$(y_{i,t+h} - \hat{y}_{i,t+h|\tau}) = \alpha_{i,h} + \beta_{i,h} \Delta \hat{y}_{i,t+h|\tau} + e_{i,t+h|\tau} \quad (1)$$

In equation (1), the forecasted macroeconomic variable  $y_i$  (annualized quarterly real GDP growth or annualized quarterly core PCE inflation) for calendar quarter  $t$   $h$  quarters ahead is  $y_{i,t+h}$ , the forecast of  $y_{i,t+h}$  on day  $\tau$  is  $\hat{y}_{i,t+h|\tau}$ ,  $\Delta$  is the first difference operator, and  $e$  is the model error. We estimated equation (1) with ordinary least squares (OLS) for each macroeconomic variable-forecast horizon, which implies that we evaluated the staff's forecasts using root mean squared error. Efficient forecasts imply that both  $\alpha$  and  $\beta$  are zero. A positive estimated  $\alpha$  indicates that the staff's average forecast was too low (in percentage points, not percent) and a positive estimated  $\beta$  indicates that the staff, on average, should have revised its forecast more than it actually did (by  $\beta \times 100\%$ , not percentage points).

We estimated equation (1) from 2001 to 2011 separately using: (1) only Greenbooks, and (2) using all forecasts in our high-frequency forecast dataset, which also included Greenbooks. Although Greenbooks extend back to 1967, our high-frequency forecast dataset has between-Greenbook observations only since 2001, and 2011 was the latest Greenbook that complied with the FOMC information security situation as of June 1, 2017, which is approximately when we registered our preanalysis plan and finished collecting our data.<sup>15</sup>

Forecast horizons,  $h$ , were from one-quarter backcasts through two quarter ahead forecasts for GDP, and one-quarter backcasts through one-quarter ahead forecasts for inflation. We chose these forecast horizons because they were the ones available in our high-frequency dataset with reasonable sample sizes, though the one-quarter ahead inflation regressions still have a somewhat small sample.

Because the staff's forecasts are tied to FOMC meetings, in our high-frequency forecast dataset we indexed forecast horizons based on when the forecasts were made as was relative at the next regularly scheduled FOMC meeting. This timing convention implied that, around the times that calendar quarters changed,  $\Delta \hat{y}_{i,t+h|\tau}$  was a forecast of  $y_{i,t}$  made  $h$  calendar

---

<sup>15</sup>Greenbooks are publicly released with an approximately six-year lag.

quarters ahead minus a forecast of  $y_{i,t}$  made  $h+1$  calendar quarters ahead. For example, if we observed a current-quarter forecast of GDP in April (which would be a forecast of Q2) that was before an April FOMC meeting, and the last observed forecast of Q2 was in March after a March FOMC meeting, then  $\Delta\hat{y}_{i,t+h|\tau}$  for the current-quarter GDP specification ( $h = 0$ ) was the current-calendar-quarter forecast of GDP in April minus the one-calendar-quarter-ahead forecast of GDP in March. This timing convention also implied that, because we needed at least two forecasts to compute  $\Delta\hat{y}_{i,t+h|\tau}$ , for our high-frequency regressions we dropped Greenbook forecasts of  $y_{i,t+h}$  where we did not observe at least one between-Greenbook forecast of  $y_{i,t+h}$ .

For actual values of GDP and inflation,  $y_{i,t}$ , we used the BEA's third-release estimates.<sup>16</sup> As there was a separate regression of equation (1) for each macroeconomic variable-horizon-dataset combination, there were fourteen regressions of equation (1).<sup>17</sup>

For the regressions using only Greenbooks, we estimated an unweighted equation (1). For regressions using our high-frequency dataset, we weighted equation (1) by the number of calendar days between forecasts.<sup>18</sup>

We inter-temporally aggregated forecast observations so that the minimum absolute value of the unweighted  $\Delta\hat{y}_{i,t+h|\tau}$  was one basis point. This procedure dropped forecasts that were equivalent to the previously observed forecasts, and it also dropped other forecasts where the forecasts did not materially revise. This aggregation was to avoid potential downward bias in  $\beta$ , as the staff updates its forecasts in batch mode but our dataset may have recorded

---

<sup>16</sup>While the vintage of data used for actual values can affect inferences about the underlying forecasts (Koenig, Dolmas & Piger (2003)) it was not obvious to us at the time we composed our preanalysis plan (and is still not obvious to us) that one vintage of real-time data would be superior to others. See Croushore (2011) for a review of the real-time data literature.

<sup>17</sup>Two datasets (Greenbook-only and high-frequency) with four horizons for GDP (eight regressions) plus the two datasets with three horizons for inflation (six regressions).

<sup>18</sup>We chose this weighting scheme for regressions that used our high-frequency dataset because our high-frequency dataset is a mixed-frequency dataset that has a different number of forecasts between FOMC meetings. The mixed-frequency nature of the data is due to both the staff's changing forecast revision methods and changing recording standards. We wished to give forecast revisions made in response to when the staff had accumulated relatively more information a higher weight and we proxied information accumulation with the number of calendar days between forecasts. For Greenbook regressions, since Greenbooks are about evenly spaced throughout a year, we thought that there was no need to weight.

forecasts at a higher frequency than the frequency that the staff updated.<sup>19</sup>

Table 2 shows our results from estimating equation (1) using real GDP growth forecasts. Each column is a separate regression by forecast horizon, increasing in the horizon from left to right. The top panel shows results from our high-frequency dataset and the bottom panel shows results from using Greenbooks only.

Broadly speaking, the high-frequency regressions show similar results to the Greenbook-only regressions. The tests for coefficient equality between high-frequency and Greenbook regressions do not show statistically significant differences, as shown in the last two rows of Table 2.

There is some evidence that, on average, the staff’s current-quarter GDP forecasts *underrevised*. The coefficient of 0.52 in the current-quarter GDP regression using our high-frequency dataset indicates that the optimal revision would have been for the staff to revise their current-quarter GDP forecasts by 52 percent *more* than the staff actually did ( $p = 0.08$ ).<sup>20</sup> For example, if the staff received better-than-expected news that led it to increase its current-quarter GDP forecast, then our result suggests that the staff should have increased its current-quarter GDP forecast by an additional 52 percent.

Table 3 shows equation (1) results using core PCE inflation forecasts. Broadly speaking the high-frequency and Greenbook-only regressions still show similar results. There is a bit of evidence that the staff’s inflation backcasts overrevised ( $p = 0.04$ ), and the one-quarter-ahead regression shows inefficiencies. But because our sample of one-quarter-ahead inflation forecasts is small, we do not want to push these inflation results too hard, and we believe our inflation results show weaker evidence of inefficiency than our GDP results.<sup>21</sup>

---

<sup>19</sup>This potential downward bias in  $\beta$  is analogous to computing abnormal returns from high-frequency stock price data and running into “thin-trading” problems, where shares are not traded frequently enough to cause changes in prices commensurate with shifts in demand or supply and, as such, causes bias towards zero in the estimated  $\beta$  of a regression of a portfolio return on a market return. For a further description of “thin-trading” see Sercu, Vandebroek & Vinaimont (2008).

<sup>20</sup>p-values in this section refer to two-sided hypothesis tests where the null is that the referenced coefficient(s) equal zero.

<sup>21</sup>We also checked all of our results by re-running prespecified regressions without the largest 1 percent of residuals in magnitude, rounded down, which did not materially affect the results. Regressions without outliers are in appendix B, Tables 12 and 13.

### 3.2 Time-varying Efficiency — Main Specification

Our second set of regression results were augmented Mincer & Zarnowitz (1969) style regressions that tested for time-varying forecast efficiency. In our preanalysis plan, we designated these regressions as our main specifications.<sup>22</sup> The regressions were of the form:

$$(y_{i,t+h} - \hat{y}_{i,t+h|\tau}) = \alpha_{i,h} + \beta_{i,h}\Delta\hat{y}_{i,t+h|\tau} + \gamma_{i,h}I(\tau) + \lambda_{i,h}I(\tau)\Delta\hat{y}_{i,t+h|\tau} + e_{i,t+h|\tau} \quad (2)$$

where  $I(\tau)$  is an indicator variable for when  $\tau$  is within 14 calendar days of the beginning of a regularly scheduled FOMC meeting, including the first day of a regularly scheduled FOMC meeting. Therefore, in equation (2),  $\alpha$  is the mean forecast error when a forecast is produced more than 2 weeks before a FOMC meeting,  $\gamma$  is the difference in mean forecast error between forecasts produced before and during the 2 week period before a FOMC meeting (such that  $\alpha + \gamma$  is the mean forecast error when the forecast is produced with fewer than 2 weeks until a FOMC meeting),  $\beta$  is the mean relationship between the staff's forecast revisions and its forecast errors for forecasts made at least 14 calendar days before a regularly scheduled FOMC meeting, and the sum  $\beta + \lambda$  is the mean relationship between the staff's forecast revisions and its forecast errors for forecasts made fewer than 14 calendar days prior to a regularly scheduled FOMC meeting. The estimator, sample restrictions, weights, and data transformations were the same as in subsection 3.1.

Testing for time-varying biases and time-varying propensities for the staff to revise forecasts using a cutoff date of 14 days until a FOMC meeting through our dummy  $I(\tau)$  was somewhat arbitrary, though it was based on our understanding of the production schedule for Greenbooks. That said, we could have defined  $I(\tau)$  as say, 15 calendar days to include the second day of a regularly scheduled FOMC meeting, or to include up to three (instead of two) weeks before a FOMC meeting, or any number of other cutoffs, and just reported the

---

<sup>22</sup>Designating a set of main specifications in the preanalysis plan follows the recommendation of Casey, Glennerster & Miguel (2012).

specification that gave us “better” (read: more publishable) results. But by defining  $I(\tau)$  in our preanalysis plan, we can assure you that you are viewing the entirety of our results and that we do not have pages and pages of unreported regressions tucked away in file drawers.

This paper is the first paper that is able to estimate equation (2) because our high-frequency dataset collects staff forecasts made more than two weeks from a FOMC meeting, whereas the staff always creates Greenbooks within two weeks of a FOMC meeting.

Table 4 shows our results from equation (2) for real GDP growth forecasts, which show some evidence of time-varying forecast inefficiency.

Focusing first on the staff’s forecast revisions, there is some evidence that the staff suboptimally revises its real GDP growth forecasts when it is more than two weeks from a regularly scheduled FOMC meeting. For example, in the second column the point estimate of 0.63 on the staff’s forecast revision indicates that the staff’s current-quarter GDP forecasts that were made more than two weeks from a FOMC meeting *underrevised*, and should have revised by 63 percent more than they actually did ( $p = 0.09$ ). Furthermore, in the fourth column, the point estimate of -0.55 indicates that the staff’s two-quarter ahead real GDP forecasts that are made more than two weeks from a FOMC meeting *overrevised*, and should have revised 55 percent less than they actually did ( $p = 0.08$ ). That said, we were unable to reject the null hypothesis of efficient revisions for GDP forecasts made within two weeks of a FOMC meeting for any forecast horizon, as shown by the fact that  $Revision + Revision \times I(\tau)$  is not statistically significant (the last row of Table 4, minimum  $p = 0.40$  for two-quarter ahead forecasts).

Table 4 has some stronger evidence of time-varying biases in the staff’s real GDP growth forecasts, although the economic magnitude of the biases is small. Current-quarter GDP forecasts made within 14 days of a FOMC meeting were too low by about a quarter percentage point ( $p = 0.01$ ), and two-quarter ahead forecasts were too high by about half a percentage point ( $p = 0.01$ ).

Table 5 shows our time-varying efficiency results for core PCE inflation, which show lim-

ited evidence for inefficiency. The strongest evidence for inefficiency is for inflation backcasts made within 14 days from the start of a regularly scheduled FOMC meeting, where the table suggests that the staff *overrevises* by an average of 49 percent ( $Revision + Revision \times I(\tau) = -0.49$ ,  $p = 0.04$ ). And while some of the point estimates on the one-quarter-ahead regressions also suggest inefficiencies, the sample size of this regression is small, so we do not want to overemphasize the one quarter ahead inflation results.<sup>23</sup>

### 3.3 What Information Could the Staff Use to Improve its Forecasts?

So far we showed that (1) on average, forecasts from our high-frequency dataset have similar properties to Greenbooks, (2) evidence that both our high-frequency dataset GDP forecasts and Greenbook GDP forecasts have some inefficiencies, (3) some weaker evidence of inefficient inflation forecasts, and (4) inefficiencies may be time-varying between FOMC meetings (as opposed to say, across decades or pre/post Great Moderation),<sup>24</sup> which is a new finding that is only possible using our high-frequency forecast dataset.

We now examine whether the staff could have used high-frequency information from financial markets to better construct its forecasts.

#### 3.3.1 A High-frequency Measure of the Market’s Reaction to Macroeconomic News

We begin by calculating a high-frequency measure of how the market interprets macroeconomic news: the summation S&P 500 return-weighted standardized values for a sequence of economic data releases that Bloomberg forecasts (Bloomberg Finance LP 2017):

$$news_{\tau} = \sum_{s=0}^{70} \sum_{i=1}^I \left( r_{i,\tau} \times \frac{|y_{i,\tau} - \tilde{y}_{i,\tau-1}|}{\tilde{\sigma}_{i\tau}} \right)_{\tau-s} \quad (3)$$

<sup>23</sup>Removal of outliers did not materially affect these results. See the appendix B, Tables 14 and 15.

<sup>24</sup>For example, Tulip (2009) analyzes Greenbooks before and after the Great Moderation.

where  $y_{i,\tau}$  denotes the value of a macroeconomic data release for variable  $y_i$  that is released on calendar day  $\tau$ ,  $\tilde{y}_{i,\tau-1}$  is the eve-of-release Bloomberg median forecast of  $y_{i,\tau}$ , and  $r_{i,\tau}$  is the percent change of the S&P 500 futures index from 5 minutes before to 25 minutes after the release of  $y_{i,\tau}$ .<sup>25</sup> There may be multiple pieces of news on the same day, or even at the same time. Furthermore, the news contained in release  $y_i$  occurs on day  $\tau$ , no matter the reference date of the release. Since macroeconomic data are released with a lag, the news on day  $\tau$  may pertain to economic activity from a previous month or quarter. The rolling two-year standard deviation of median Bloomberg forecast errors for  $y_i$ , denoted  $\tilde{\sigma}_{i,\tau}$ , standardizes the news content of each release  $y_{i,\tau}$ .<sup>26</sup> Lastly, we use the absolute value of the Bloomberg forecast error,  $|y_{i,\tau} - \tilde{y}_{i,\tau-1}|$ , so that it is the market reaction to  $y_{i,\tau}$ , as contained in  $r_{i,\tau}$ , that determines whether an individual data release is positive or negative news.

Much like defining  $I(\tau)$ , we had to make somewhat arbitrary choices to calculate a single summary measure of how the market interprets macroeconomic news, though our formulation is similar to that used by Scotti (2016). Certainly our measure is imperfect and there are other means to measure the market’s interpretations. But at the time of writing our preanalysis plan, we did have an alternative that strictly dominated  $news_\tau$ .

One omission from our preanalysis plan was which Bloomberg-forecasted series go into  $news_\tau$ . When we wrote our preanalysis plan, we thought that we were using all data from Bloomberg—that turned out not to be the case. We found out after estimating and viewing our models that we had access to only a subset of macroeconomic data from Bloomberg, though we used everything that we had access to in June 2017 to create  $news_\tau$ . The 62 series that make up our news index are listed in appendix B, tables 18 and 19, which cover most important U.S. macro data. In keeping with our commitment to avoid p-hacking, we have only used the calculated  $news_\tau$  with our original data and have not attempted to manipulate what data make up  $news_\tau$ . Therefore, we hope that you continue to find assurance that,

---

<sup>25</sup>For example, for the employment report that is released at 8:30 AM eastern time,  $r_{i,\tau}$  is the percent change of the S&P 500 futures index from 8:25 AM to 8:55 AM eastern time.

<sup>26</sup>We calculate  $\tilde{\sigma}_{i,\tau}$  excluding the forecast miss on day  $\tau$ .

because we defined  $news_\tau$  in our preanalysis plan before viewing our results and, at least, have not changed what data compose  $news_\tau$  after we viewed our model results, you are seeing unbiased estimates of the predictive power of market’s interpretations of macroeconomic news on the staff’s forecast errors.

Figure 4 plots  $news_\tau$ . Our measure suggests that, on average and perhaps reassuringly, the market is unsurprised by news as the mean of  $news_\tau$  is about zero. However, the variance of  $news_\tau$  spikes during the Great Recession and remains elevated thereafter, indicating that the market was more surprised by data —both positively and negatively —during and after the Great Recession. Interestingly,  $news_\tau$  rises at the beginning of the 2001 and 2008 recessions.

### 3.3.2 Predicting the Staff’s Forecast Errors Using the Market’s Reaction to Macroeconomic News

We used our measure of the market’s reaction to macroeconomic news,  $news_\tau$ , to attempt to predict the staff’s forecast errors using our high-frequency dataset. Our regressions that used  $news_\tau$  took the form:

$$(y_{i,t+h} - \hat{y}_{i,t+h|\tau}) = \alpha_{i,h} + \beta_{i,h}\Delta\hat{y}_{i,t+h|\tau} + \gamma_{i,h}I(\tau) + \lambda_{i,h}I(\tau)\Delta\hat{y}_{i,t+h|\tau} + \eta_{i,h}news_\tau + \theta_{i,h}I(\tau)news_\tau + e_{i,t+h|\tau} \quad (4)$$

where, as in subsections 3.1 and 3.2, the forecasted macroeconomic variable  $y_i$  for calendar quarter  $t$   $h$  quarters ahead is  $y_{i,t+h}$ , the staff forecast of  $y_{i,t+h}$  on day  $\tau$  is  $\hat{y}_{i,t+h|\tau}$ ,  $\Delta$  is the first difference operator,  $I(\tau)$  is an indicator variable when  $\tau$  is within 14 calendar days from the start of a regularly scheduled FOMC meeting,  $e$  is the model error, and  $news_\tau$  is our measure of the market’s reaction to macroeconomic news in equation (3). Sample selection, data transformations, the estimator, and so on are the same as in subsections 3.1 and 3.2. In equation (4),  $\eta$  is the average relationship between our measure of the market’s reaction to macroeconomic news and the staff’s forecast errors for forecasts made at least 14 days

from the start of a regularly scheduled FOMC meeting, and  $\eta + \theta$  is the same relationship for forecasts that are fewer than 14 days from the start of a meeting.

Table 6 shows our estimation results of equation (4) using real GDP growth forecasts. The staff’s GDP forecast errors for backcasts up to one-quarter-ahead forecasts are negatively correlated with  $news_\tau$ , which indicates that better-than-expected macroeconomic news is correlated with more accurate staff forecasts ( $p < 0.01$  for current-quarter forecasts). This negative correlation holds for both GDP forecasts made more than two weeks from a FOMC meeting and for GDP forecasts made within two weeks of a FOMC meeting for up through one-quarter-ahead forecasts.

In terms of effect sizes, a one standard deviation increase in  $news_\tau$  implies that the staff’s GDP current-quarter errors are, on average, smaller by about 0.4 percentage points.<sup>27</sup> This finding is not the same as finding that the staff is more accurate in expansions than in recessions, as  $news_\tau$  experiences local peaks and troughs both in and outside of recessions.

Table 7 shows the results for the correlations between  $news_\tau$  and core PCE inflation forecast errors. Unlike the results for GDP,  $news_\tau$  does not predict the staff’s inflation forecast errors, which gives some additional evidence that the staff forms its inflation projections efficiently.<sup>28</sup>

## 4 Erroneous Interpretations Under “Cherry Picking”

Our econometric analysis was disciplined by a preanalysis plan. Preanalysis plans for randomized control trials are prominent outside of economics, and within economics there is some evidence that they are becoming more common. For example, the number of registered preanalysis plans on the American Economic Association’s registry for randomized control trials has trended up since 2013 (Vilhuber, Turitto & Welch (2020), Figures 13

---

<sup>27</sup>The standard deviation of  $news_\tau$  is about 7, the coefficient on  $news_\tau$  for GDP current-quarter forecasts is -0.06, so  $7 \times -0.06 \approx -0.4$ .

<sup>28</sup>Removal of outliers, shown in appendix B, Tables 16 and 17, do not change the conclusions for either GDP or inflation.

and 14). Casey, Glennerster & Miguel (2012) provide an excellent discussion of the value of preanalysis plans for randomized control trials, with a focus on evaluating institutional outcomes.

At the same time, preanalysis plans are useful for observational studies, including this study. Data mining or “cherry picking” is a risk for observational studies because of the infinite number of assumptions researchers can make when no dominant assumption exists.<sup>29</sup> This infinite number of potential assumptions implies an infinite number of potential models and researchers cannot test every model. Typically a study will select a baseline model. The study will then conduct a finite number of robustness checks as reasonable deviations from the baseline model, usually by perturbing one assumption at a time, and then report a subset of those robustness checks. The gap between the infinite number of potential models and the finite number of reported models can lead to “cherry picking” and publication bias.<sup>30</sup>

We thought that a preanalysis plan for this study would be particularly useful as we were both involved in creating the Greenbooks (though we were not involved with the particular forecasts in our high-frequency forecast dataset). Our involvement with creating the Greenbooks meant that we are subject to certain biases, possibly in favor of finding that the Greenbooks are efficient. In the absence of a preanalysis plan, these biases could have lead us to, consciously or unconsciously, make *ex post* assumptions and selectively “cherry pick” results. We used a preanalysis plan to stomp out these biases, to select assumptions *ex ante*, and to report, to the extent possible, unbiased results.

To demonstrate the value of a preanalysis plan for disciplining observational research, we p-hacked a set of results using GDP forecasts and our main specification, equation 2, that give a different interpretation of the staff’s propensity to revise its forecasts than our preanalysis plan results. We adjusted three assumptions —the weights, the threshold for

---

<sup>29</sup>Vivalt (2019) and Brodeur, Cook & Heyes (Forthcoming) find evidence “cherry picking” is a bigger problem for observational studies than for randomized control trials.

<sup>30</sup>Evidence that publication bias exists includes: Brodeur, Lé, Sangnier & Zylberberg (2016), Chang & Li (2018), Vivalt (2019), Blanco-Perez & Brodeur (2020), and Brodeur, Cook & Heyes (Forthcoming). Chen & Zimmermann (2020) provide some potential limits on publication bias in the asset pricing literature.

keeping a forecast revision, and the definition of  $I(\tau)$  —in ways we think are inferior to our preanalysis plan assumptions and, at the same time, in ways that other researchers could plausibly think have similar academic rigor to our preanalysis plan assumptions (or at least are not vastly inferior to our preanalysis plan assumptions) *ex post*. To mimic the refereeing process, we selected a baseline model and performed “cherry picked” robustness checks to the baseline model by perturbing the adjusted assumptions, one at a time, in ways that could also be justified by other researchers as reasonable (or at least not horrific) *ex post*.

Table 8 shows our “cherry picked” baseline where we estimate unweighted regressions, increase the threshold for keeping a forecast revision to be at least 0.1 percentage points in magnitude, and define  $I(\tau)$  as 28 days.<sup>31</sup> While our main prespecified results in Table 4 show that the staff’s current-quarter forecasts *underrevise*, the “cherry picked” baseline in Table 8 indicates the staff’s backcasts and one-quarter ahead forecasts *overrevise*, and there is no evidence for the current-quarter forecasts underrevising.

Tables 9 - 11 show robustness checks to the “cherry picked” baseline. Sequentially, the tables perturb the “cherry picked” baseline assumptions by: weighting by the number of weeks between forecasts (rounded up), keeping all forecast revisions, and setting  $I(\tau)$  as 21 days. These “cherry picked” robustness tests largely confirm the “cherry picked” baseline and are at odds with our prespecified results. Taken together, Tables 8 - 11 demonstrate the importance of a preanalysis plan.

## 5 Conclusion

We created, using various archived documents, a new high-frequency dataset of the Federal Reserve Board staff’s forecasts of real GDP growth and core PCE inflation. These data record the staff’s forecasts roughly weekly and complement the existing data from Greenbooks, which the staff produce eight times per year.

---

<sup>31</sup>Our preanalysis plan specified weighted regressions by the number of calendar days between forecasts, to use forecast revisions of at least 1 basis point in magnitude, and set  $I(\tau)$  as 14 days.

Using these new data and a preanalysis plan, we analyzed the efficiency of the staff's high-frequency forecasts. We found evidence of inefficiencies in the staff's GDP forecasts—current-quarter forecasts tended to underrevise. We found some weaker evidence for inefficiency of inflation forecasts. For GDP, we found between-FOMC-meeting inefficiencies. Forecasts made at least two weeks from a FOMC meeting were less efficient.

In addition to analyzing time-varying efficiencies of the staff's forecasts, we also found some evidence that a summary measure of the market's reaction to macroeconomic news predicts the staff's GDP forecast errors. This result suggests that the staff does not efficiently incorporate news from financial markets in forecasting GDP, though the same summary measure did not predict the staff's inflation forecast errors.

Although we found time-varying inefficiencies between FOMC meetings in the staff's GDP forecasts, we cannot say for certain why such inefficiencies exist, though we speculate on four possibilities.

First, though we prespecified a vintage for actual GDP and inflation (the BEA's third release) to avoid p-hacking our way results, it is possible that the third release is not what the staff was forecasting. If the staff was targeting a different vintage of data AND the revisions to data are predictable AND the predictability of revisions is tied to the FOMC meeting schedule, then we could find time-varying forecast inefficiencies between FOMC meetings when we should find none (or we could find no inefficiencies when we should find some).

Because all three of these conditions need to be satisfied in order to support our selection of the BEA's third release as the cause of our results, we view this explanation as very unlikely. Though Figure 3 shows the staff was not targeting BEA's first release, there is no other information that we are aware of on what vintage of data the staff was targeting. Furthermore, the literature on the predictability of U.S. macroeconomic data revisions is inconclusive (Faust, Rogers & Wright (2005), Aruoba (2008)). We also know of no research that ties the potential predictability of data revisions (or lack thereof) to the FOMC meeting schedule but we cannot think of a good reason why such predictability would depend on the

FOMC meeting schedule.

Second, it is possible that our prespecified evaluation criterion for forecast efficiency —minimizing root mean squared error —isn't exactly what the staff was trying to do.

Root mean squared error evaluates forecasts as unconditional. And although Figure 3 shows the staff revises its forecasts in a way that is consistent with minimizing root mean squared error, the staff's forecasts are conditional on an assumed path for the federal funds rate.<sup>32</sup> Evidence from Berge, Chang & Sinha (2019) shows that Greenbook forecasts of macroeconomic variables appropriately condition on the path for the federal funds rate.

The forecast errors of conditional forecasts contain two components —the unconditional forecast error and the forecast error of the conditioning variable. It is possible that the staff's forecast errors for the conditioning variable (the federal funds rate) are what causes our results. But for short-horizon forecasts the difference between the forecasted conditioning variable and the actual conditioning variable is small. In the extreme cases, for backcasts there is no difference and for current-quarter forecasts the conditioning variable is set for part of the quarter —so for these cases forecast errors of the conditioning variable are very unlikely to be the source our results.<sup>33</sup>

Third, it is possible that the information set of the staff varies along the same timetable as FOMC meetings and time-varying information causes our results, though we also view this explanation as unlikely.

The dates of FOMC meetings tend to be selected to coincide with when major traditional macroeconomic data (which are often monthly data) become available. As such, the information the staff has may be correlated with the timing of FOMC meetings. But to the extent that the staff can extract the same signals about the economy from non-traditional high-frequency data as the staff can from traditional data (for example, Automatic Data Processing payroll information instead of the employment report), the likelihood that the

---

<sup>32</sup>The staff assumes a path for the federal funds rate to avoid confounding its forecasts with the opinions of the Board of Governors.

<sup>33</sup>Furthermore, if this explanation were correct then the staff's forecast errors for the conditioning variable would be the ones exhibiting time-varying inefficiency between FOMC meetings.

staff’s information is a function of the timing to the next FOMC meeting is lessened.

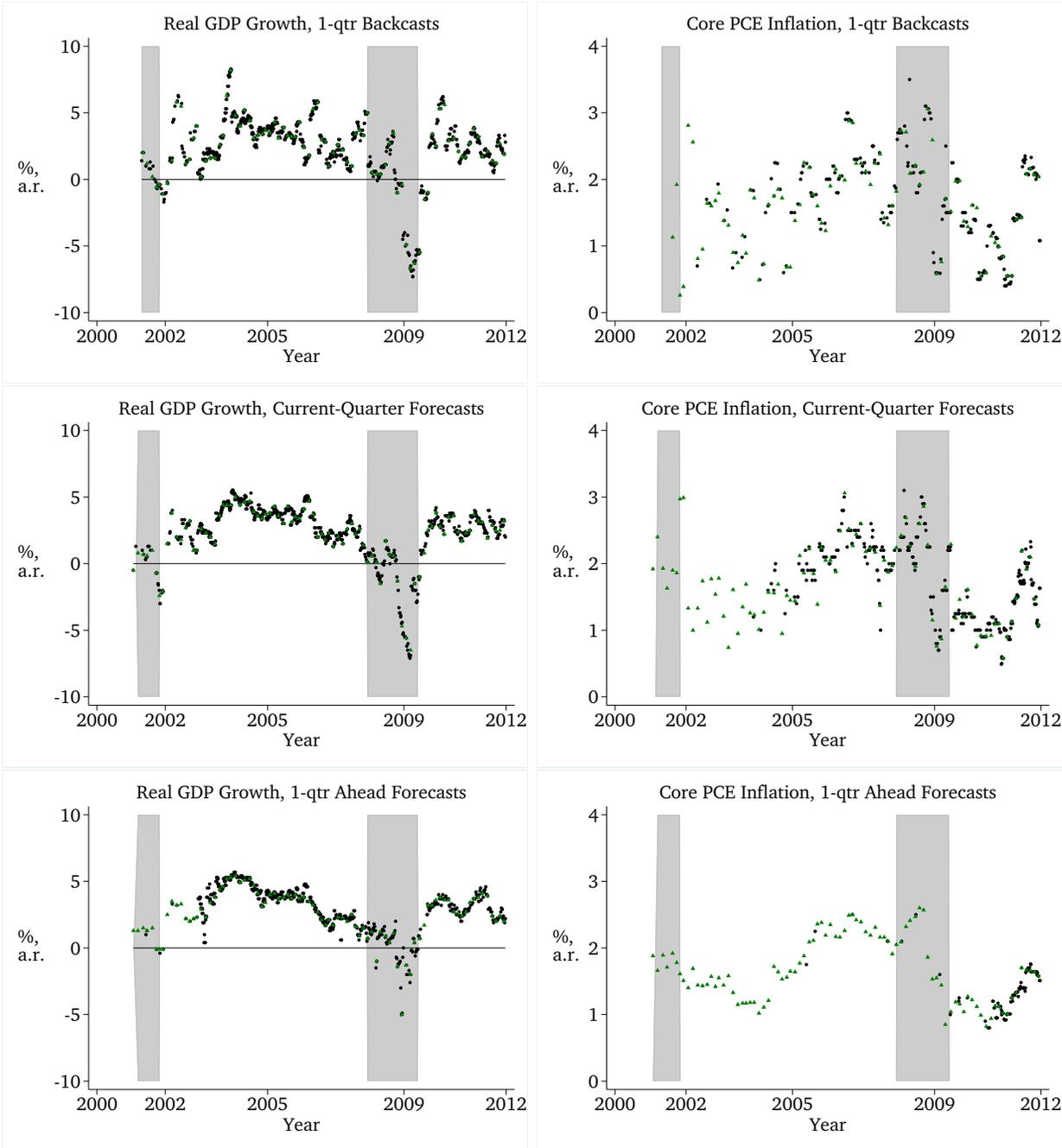
Furthermore, within a year FOMC meetings occur on different days of the month and traditional macroeconomic data are released around the same time each month. For example, the employment report is on the first Friday of each month. Because our results are averages across all FOMC meetings, the likelihood that the time-varying information is the cause of our results is lessened further.

Fourth —and in our view the most likely possibility —the staff could forecast differently in relation to the schedule of FOMC meetings. The primary purpose of the staff is to serve the American public by being an asset for the Board of Governors. To that end, the few weeks before a FOMC meeting are when the staff focuses on creating the Greenbooks, which are the forecasts that receive the most scrutiny from the staff and from the Board of Governors, as Greenbooks have the most direct role in monetary policy-making. Although the staff continues to monitor the economy away from FOMC meetings, the staff puts a larger focus on longer-run productivity-improving projects, such as infrastructure maintenance, away from meetings. It may be optimal for the staff to be less attentive to between-Greenbook forecasts in order to maximize its long-run usefulness for the Board of Governors.

## **6 Acknowledgments**

We completed most of this paper while Levinson was a Senior Research Assistant at the Board. We thank Katherine Arnold, Travis J. Berge, Garret Christensen, Christopher Karlsten, Ekaterina Peneva, Jeremy B. Rudd, Steven A. Sharpe, Nitish R. Sinha, Stacey Tevlin, and David Wilcox for helpful comments. We thank Amanda G. Bauer, Kelsey O’Flaherty, and Sasha Ruby for research assistance. We are responsible for any errors.

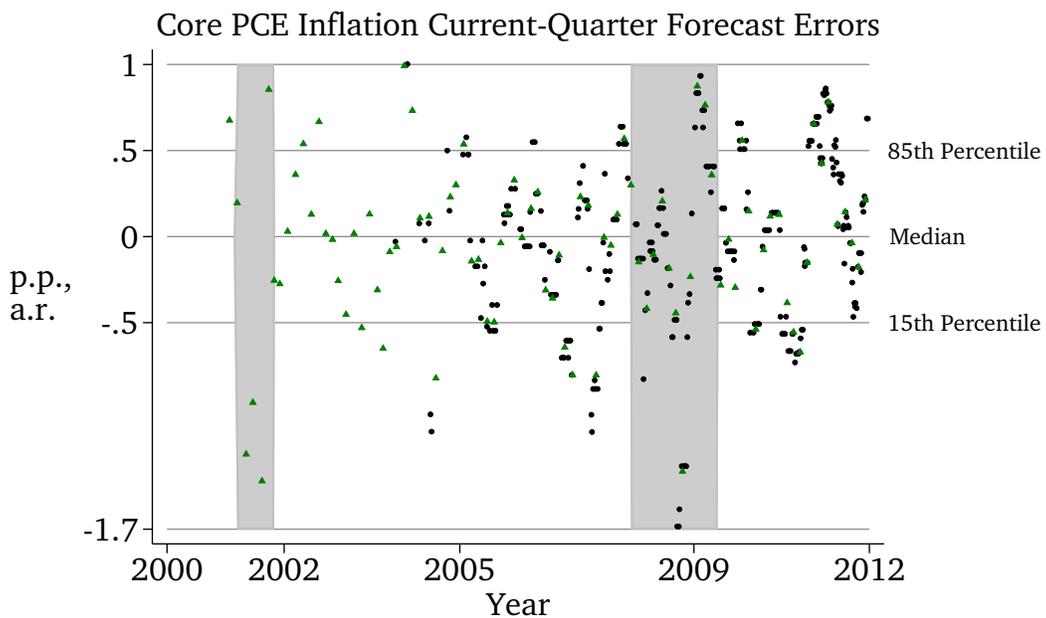
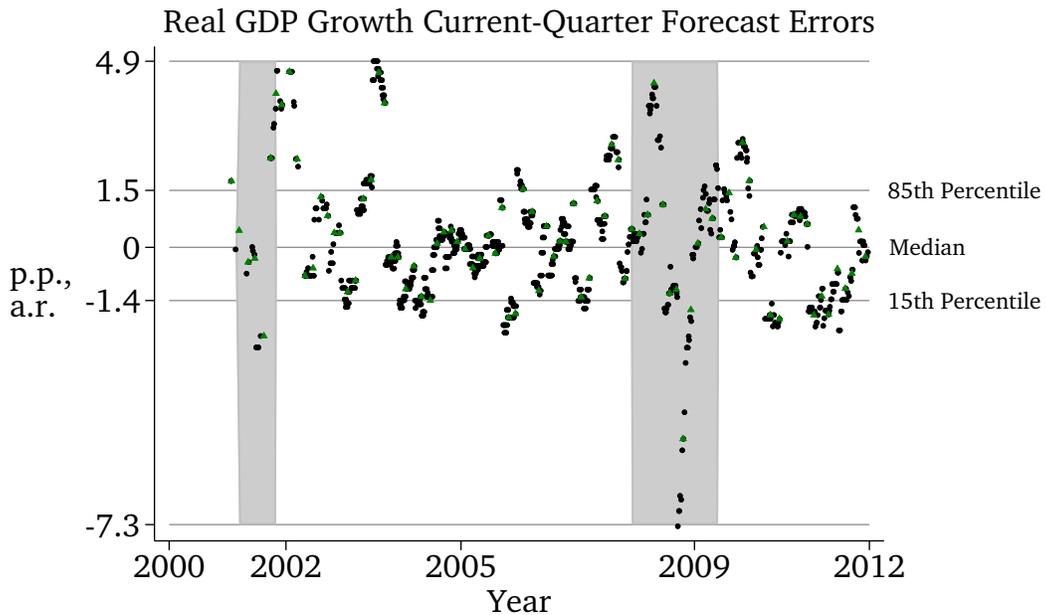
Figure 1: Near-term Board Staff Forecasts of Real GDP Growth and Core PCE Inflation.



**Description:** High-frequency staff forecasts of quarterly annualized real GDP growth (left) and core PCE inflation (right). Black dots denote between-Greenbook forecasts. Green triangles denote Greenbook forecasts. NBER recessions shaded. Year marks at January 1 of each year.

**Interpretation:** The staff updates its forecasts of GDP and inflation frequently between Greenbooks. Our high-frequency forecast dataset contains an average of about 80 forecasts of GDP for each horizon per year but has fewer forecasts for inflation.

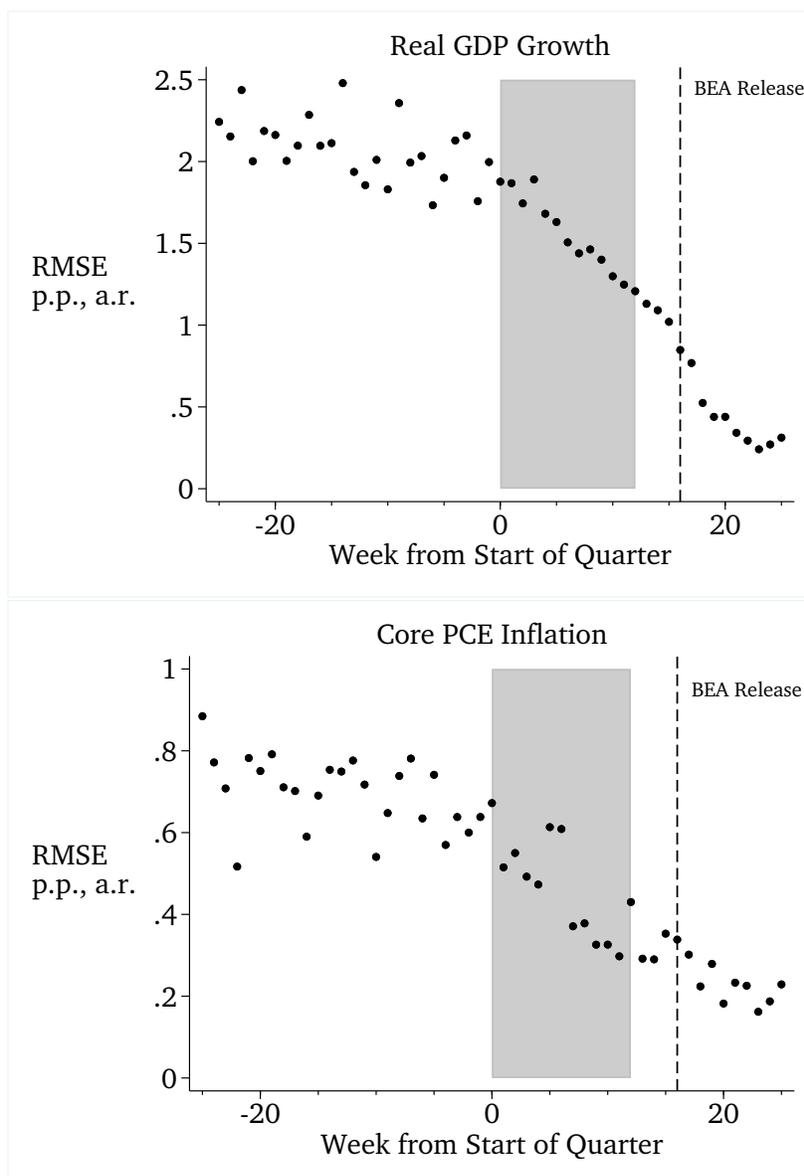
Figure 2: Board Staff Current-Quarter Errors of Real GDP Growth and Core PCE Inflation.



**Description:** High-frequency forecast errors of the current-quarter (nowcasts) for quarterly annualized real GDP growth (top) and core PCE inflation (bottom). Black dots denote between-Greenbook current-quarter errors. Green triangles denote Greenbook errors. NBER recessions shaded. Year marks at January 1 of each year. Error is the BEA's third release minus the forecast.

**Interpretation:** The current-quarter errors for both GDP and inflation are about unbiased, even during the Great Recession, and the middle 70 percent of errors is symmetric around zero. But the staff exhibit large errors even outside of the 2001 and 2008 recessions.

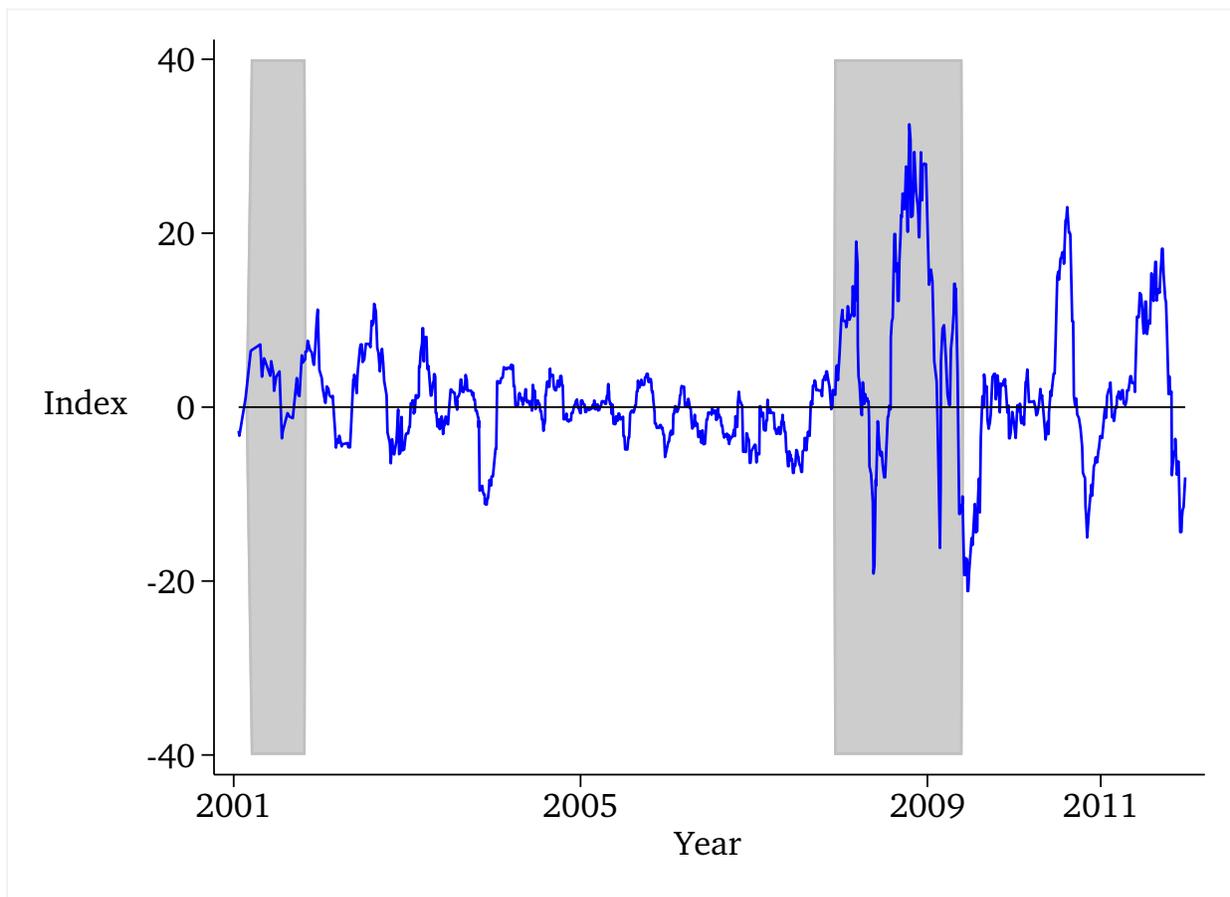
Figure 3: High-frequency Performance of Board Staff Forecasts of Real GDP Growth and Core PCE Inflation.



**Description:** Figure shows the root mean squared forecast errors of staff projections of annualized quarterly real GDP growth (top) and core PCE inflation (bottom) by the week to the start of the quarter. Actual data are the BEA’s third-release estimates. Week zero denotes the start of the quarter forecasted—for example, the first week of January when the quarter forecasted is Q1. Approximately, negative weeks are *forecasts*, weeks 0-12 are current-quarter forecasts (*nowcasts*, shaded gray), and weeks greater than 12 are *backcasts* in calendar time. Vertical dashed line indicates the approximate release of the BEA’s first-release estimate.

**Interpretation:** The staff continuously updates and improves its forecasts of GDP and inflation between Greenbooks. The staff’s root mean squared error falls considerably when current-quarter data become available and continues to fall even after the BEA releases its first estimate of GDP and inflation.

Figure 4: An Index of the Market's Reaction to Macroeconomic News



**Description:** The index is calculated as a rolling sum of S&P 500 futures returns weighted by the size of normalized forecast errors from Bloomberg Finance LP (2017). NBER recessions shaded. Year marks at January 1 of each year. Higher values of this index indicate that the market interpreted incoming macroeconomic news as better than expected, relative to its eve-of-release expectations. See text in sub-subsection 3.3.1 for further details.

**Interpretation:** Our measure of news is mean stationary, though the variance spikes during the Great Recession and remains elevated thereafter, indicating that the market was more surprised by data during and after the Great Recession, relative to the period before. Interestingly, the measure rises at the beginning of the 2001 and 2008 recessions.

Table 1: Summary Statistics of High-frequency Forecast Errors.

Forecast Horizon	-1	0	1	2
<b>GDP</b>				
Mean	.08	.11	-.29	-.63
Std. Dev.	.79	1.62	2.01	2.1
Min.	-2.34	-7.34	-7.49	-7.29
Max.	3.8	4.9	6.53	5.45
N. Obs.	1028	992	812	755
<b>Inflation</b>				
Mean	0	-.03	.17	.3
Std. Dev.	.29	.49	.62	.62
Min.	-1.44	-1.68	-1.68	-1.64
Max.	.68	1	1.34	1.36
N. Obs.	336	398	183	154

**Description:** Table shows summary statistics for forecast errors of the staff projections of annualized quarterly real GDP growth and annualized quarterly core PCE inflation by forecast horizon. Error defined as the BEA's third release less the forecast.

**Interpretation:** The staff made sizable forecast errors for both GDP and inflation. The staff's forecasts are about unbiased for backcasts and current-quarter forecasts. There is some evidence of bias for further ahead forecasts, though the sample size for further ahead forecasts is smaller.

Table 2: Real GDP Efficiency Regressions Suggest Some Forecast Inefficiencies

Forecast Horizon	-1	0	1	2
	High-frequency			
Revision	-0.10 (0.11)	0.52 (0.30)	-0.32 (0.35)	-0.12 (0.25)
Constant	0.04 (0.04)	0.17 (0.09)	-0.37 (0.14)	-0.71 (0.14)
$N$	616	579	495	381
adj. $R^2$	0.00	0.01	0.00	-0.00
	Greenbook Only			
Revision	0.03 (0.07)	0.39 (0.19)	-0.29 (0.33)	-0.33 (0.40)
Constant	0.04 (0.07)	0.35 (0.18)	-0.22 (0.26)	-0.72 (0.24)
$N$	86	87	87	87
adj. $R^2$	-0.01	0.04	-0.00	-0.00
p-values:				
$H_0$ : High-Frequency Revision = Greenbook Revision	0.35	0.73	0.96	0.67
$H_0$ : High-Frequency Constant = Greenbook Constant	0.97	0.35	0.61	0.96

**Description:** Table shows estimated coefficients from the regression  $y_{i,t+h} - \hat{y}_{i,t+h|\tau} = \alpha_{i,h} + \beta_{i,h}\Delta\hat{y}_{i,t+h|\tau} + e_{i,t+h|\tau}$  for Federal Reserve Board staff projections of annualized quarterly real GDP growth by forecast horizon. The top panel uses our high-frequency forecast dataset and the bottom panel uses Greenbooks only. For Greenbook regressions, OLS standard errors in parentheses. We weight the high-frequency regressions by number of days between forecast revisions in  $\Delta\hat{y}_{i,t+h|\tau}$ , with Huber-White (White 1980) standard errors in parentheses. Hypothesis tests are two sided. Statistical significance asterisks omitted.

**Interpretation:** The regressions from both our high-frequency forecast dataset and the Greenbooks suggest, on average, some forecast inefficiencies. The current-quarter staff GDP forecasts tend to underrevise, and there is also some evidence of bias.

Table 3: Inflation Efficiency Regressions Suggest Limited Inflation Forecast Inefficiencies

Forecast Horizon	-1	0	1
High-frequency			
Revision	-0.45 (0.21)	0.23 (0.25)	-1.82 (0.82)
Constant	-0.02 (0.03)	-0.00 (0.04)	0.26 (0.10)
<i>N</i>	166	206	48
adj. <i>R</i> <sup>2</sup>	0.08	0.00	0.12
Greenbook Only			
Revision	-0.14 (0.11)	-0.18 (0.14)	0.13 (0.38)
Constant	0.01 (0.03)	-0.05 (0.05)	-0.05 (0.06)
<i>N</i>	84	87	87
adj. <i>R</i> <sup>2</sup>	0.01	0.01	-0.01
p-values:			
$H_0$ : High-Frequency Revision = Greenbook Revision	0.18	0.14	0.03
$H_0$ : High-Frequency Constant = Greenbook Constant	0.46	0.51	0.01

**Description:** Table shows estimated coefficients from the regression  $y_{i,t+h} - \hat{y}_{i,t+h|\tau} = \alpha_{i,h} + \beta_{i,h}\Delta\hat{y}_{i,t+h|\tau} + e_{i,t+h|\tau}$  for Federal Reserve Board staff projections of annualized quarterly core PCE inflation by forecast horizon. The top panel uses our high-frequency forecast dataset and the bottom panel uses Greenbooks only. For Greenbook regressions, OLS standard errors in parentheses. We weight the high-frequency regressions by number of days between forecast revisions in  $\Delta\hat{y}_{i,t+h|\tau}$ , with Huber-White (White 1980) standard errors in parentheses. Hypothesis tests are two sided. Statistical significance asterisks omitted.

**Interpretation:** The high-frequency regressions indicate some tendency, on average, for one-quarter backcasts of inflation to overrevise, but otherwise we find little evidence of inefficient inflation forecasts. The one-quarter-ahead regressions have a small sample size, so we do not want to overemphasize their results.

Table 4: Time-Varying Real GDP Efficiency Regressions Also Suggest Some Forecast Inefficiencies

Forecast Horizon	-1	0	1	2
$I(\tau)$	-0.05 (0.08)	0.18 (0.16)	0.45 (0.36)	0.39 (0.27)
Revision	-0.15 (0.14)	0.63 (0.38)	-0.22 (0.45)	-0.55 (0.31)
$I(\tau) \times \text{Revision}$	0.16 (0.26)	-0.30 (0.61)	-0.13 (0.68)	0.84 (0.47)
Constant	0.06 (0.06)	0.10 (0.12)	-0.51 (0.14)	-0.87 (0.19)
$N$	616	579	495	381
adj. $R^2$	-0.00	0.01	0.01	0.00
p-values:				
$H_0 : \text{Constant} + I(\tau) = 0$	0.88	0.01	0.86	0.01
$H_0 : \text{Revision} + I(\tau) \times \text{Revision} = 0$	0.99	0.49	0.49	0.40

**Description:** Table shows estimated coefficients from the regression  $y_{i,t+h} - \hat{y}_{i,t+h|\tau} = \alpha_{i,h} + \beta_{i,h}\Delta\hat{y}_{i,t+h|\tau} + \gamma_{i,h}I(\tau) + \lambda_{i,h}I(\tau)\Delta\hat{y}_{i,t+h|\tau} + e_{i,t+h|\tau}$  for Federal Reserve Board staff projections of annualized quarterly real GDP growth by forecast horizon, where  $I(\tau)$  is an indicator for a forecast made within 14 calendar days from the start of a regularly-scheduled FOMC meeting. We weight these regressions by number of days between forecast revisions in  $\Delta\hat{y}_{i,t+h|\tau}$ , with Huber-White (White 1980) standard errors in parentheses. Hypothesis tests are two sided. Statistical significance asterisks omitted.

**Interpretation:** There is some evidence that Federal Reserve Board staff GDP forecasts exhibit time-varying efficiency. Forecasts made at least 14 days from the start of a regularly scheduled FOMC meeting revise inefficiently, whereas we find efficient revisions for those forecasts made within 14 days from the start of a meeting. There is some stronger evidence of bias.

Table 5: Time-Varying Inflation Efficiency Regressions Also Indicate Limited Evidence for Inefficiency

Forecast Horizon	-1	0	1
$I(\tau)$	0.03 (0.06)	-0.01 (0.08)	-0.07 (0.18)
Revision	-0.18 (0.41)	0.09 (0.26)	2.32 (1.55)
$I(\tau) \times \text{Revision}$	-0.31 (0.47)	0.38 (0.54)	-4.82 (1.62)
Constant	-0.05 (0.05)	0.00 (0.05)	0.32 (0.15)
$N$	166	206	48
adj. $R^2$	0.08	0.00	0.25
p-values:			
$H_0 : \text{Constant} + I(\tau) = 0$	0.67	0.85	0.02
$H_0 : \text{Revision} + I(\tau) \times \text{Revision} = 0$	0.04	0.32	0.00

**Description:** Table shows estimated coefficients from the regression  $y_{i,t+h} - \hat{y}_{i,t+h|\tau} = \alpha_{i,h} + \beta_{i,h} \Delta \hat{y}_{i,t+h|\tau} + \gamma_{i,h} I(\tau) + \lambda_{i,h} I(\tau) \Delta \hat{y}_{i,t+h|\tau} + e_{i,t+h|\tau}$  for Federal Reserve Board staff projections of annualized quarterly core PCE inflation by forecast horizon, where  $I(\tau)$  is an indicator for a forecast made within 14 days from the start from a regularly-scheduled FOMC meeting. We weight these regressions by number of days between forecast revisions in  $\Delta \hat{y}_{i,t+h|\tau}$ , with Huber-White (White 1980) standard errors in parentheses. Hypothesis tests are two sided. Statistical significance asterisks omitted.

**Interpretation:** There is some evidence that Federal Reserve Board staff inflation backcasts made within 14 days from the start of a regularly scheduled FOMC meeting overrevised. Otherwise, there is not much evidence of inefficient inflation forecasts. The sample size of one-quarter-ahead inflation regressions is small, so we do not place too much weight on these results.

Table 6: Better-than-Expected Macroeconomic News Predicts Federal Reserve Staff Real GDP Forecast Errors

Forecast Horizon	-1	0	1	2
$I(\tau)$	-0.05 (0.08)	0.26 (0.16)	0.51 (0.34)	0.31 (0.28)
Revision	-0.16 (0.13)	0.43 (0.39)	-0.35 (0.42)	-0.66 (0.30)
$I(\tau) \times \text{Revision}$	0.13 (0.26)	-0.42 (0.53)	-0.48 (0.65)	0.97 (0.46)
$news_\tau$	-0.02 (0.01)	-0.06 (0.02)	-0.05 (0.03)	-0.04 (0.03)
$I(\tau) \times news_\tau$	0.01 (0.01)	0.00 (0.03)	-0.02 (0.05)	0.06 (0.04)
Constant	0.07 (0.05)	0.14 (0.11)	-0.47 (0.13)	-0.81 (0.19)
$N$	616	579	495	381
adj. $R^2$	0.01	0.07	0.05	0.01
p-values:				
$H_0 : \text{Constant} + I(\tau) = 0$	0.64	0.00	0.90	0.02
$H_0 : \text{Revision} + I(\tau) \times \text{Revision} = 0$	0.89	0.97	0.10	0.37
$H_0 : news_\tau + I(\tau) \times news_\tau = 0$	0.13	0.00	0.05	0.62

**Description:** Table shows estimated coefficients from the regression  $y_{i,t+h} - \hat{y}_{i,t+h|\tau} = \alpha_{i,h} + \beta_{i,h}\Delta\hat{y}_{i,t+h|\tau} + \gamma_{i,h}I(\tau) + \lambda_{i,h}I(\tau)\Delta\hat{y}_{i,t+h|\tau} + \eta_{i,h}news_\tau + \theta_{i,h}I(\tau)news_\tau + e_{i,t+h|\tau}$  for Federal Reserve Board staff projections of annualized quarterly real GDP growth by forecast horizon, where  $I(\tau)$  is an indicator for a forecast made within 14 days from the start of a regularly-scheduled FOMC meeting, and  $news_\tau$  is our measure of the market's reaction to macroeconomic news using data from Bloomberg Finance LP (2017). We weight these regressions by number of days between forecast revisions in  $\Delta\hat{y}_{i,t+h|\tau}$ , with Huber-White (White 1980) standard errors in parentheses. Hypothesis tests are two sided. Statistical significance asterisks omitted.

**Interpretation:** We find evidence that the market's reaction to macroeconomic news predicts Federal Reserve Board staff real GDP forecast errors, which suggests that the staff is not efficiently using information from financial markets to inform its forecasts of GDP. When economic news is better than expected, the forecasts of GDP are more accurate.

Table 7: Better-than-Expected Macroeconomic News Does Not Predict Federal Reserve Staff Inflation Forecast Errors

Forecast Horizon	-1	0	1
$I(\tau)$	0.04 (0.06)	0.01 (0.08)	-0.19 (0.17)
Revision	-0.17 (0.40)	0.07 (0.26)	1.43 (1.51)
$I(\tau) \times \text{Revision}$	-0.32 (0.47)	0.36 (0.55)	-3.90 (1.62)
$news_\tau$	-0.00 (0.01)	-0.00 (0.01)	-0.04 (0.01)
$I(\tau) \times news_\tau$	-0.00 (0.01)	-0.01 (0.01)	0.03 (0.01)
Constant	-0.05 (0.05)	0.01 (0.05)	0.50 (0.13)
$N$	166	206	48
adj. $R^2$	0.08	0.01	0.40
p-values:			
$H_0 : \text{Constant} + I(\tau) = 0$	0.84	0.80	0.01
$H_0 : \text{Revision} + I(\tau) \times \text{Revision} = 0$	0.03	0.38	0.00
$H_0 : news_\tau + I(\tau) \times news_\tau = 0$	0.43	0.20	0.00

**Description:** Table shows estimated coefficients from the regression  $y_{i,t+h} - \hat{y}_{i,t+h|\tau} = \alpha_{i,h} + \beta_{i,h}\Delta\hat{y}_{i,t+h|\tau} + \gamma_{i,h}I(\tau) + \lambda_{i,h}I(\tau)\Delta\hat{y}_{i,t+h|\tau} + \eta_{i,h}news_\tau + \theta_{i,h}I(\tau)news_\tau + e_{i,t+h|\tau}$  for Federal Reserve Board staff projections of annualized quarterly core PCE inflation by forecast horizon, where  $I(\tau)$  is an indicator for a forecast made within 14 days from the start of a regularly-scheduled FOMC meeting, and  $news_\tau$  is our measure of the market's reaction to macroeconomic news using data from Bloomberg Finance LP (2017). We weight these regressions by number of days between forecast revisions in  $\Delta\hat{y}_{i,t+h|\tau}$ , with Huber-White (White 1980) standard errors in parentheses. Hypothesis tests are two sided. Statistical significance asterisks omitted.

**Interpretation:** We do not find evidence that market's reaction to macroeconomic news predicts the Federal Reserve Board staff inflation forecast errors, suggesting that the staff forms its inflation forecasts efficiently.

Table 8: Main GDP Specifications Under a “Cherry Picked” Baseline Can Show Erroneous Results

Forecast Horizon	-1	0	1	2
$I(\tau)$	-0.05 (0.09)	0.40 (0.21)	0.08 (0.27)	-0.31 (0.42)
Revision	-0.31 (0.12)	0.25 (0.45)	-1.24 (0.65)	-1.07 (1.06)
$I(\tau) \times$ Revision	0.17 (0.17)	-0.11 (0.51)	1.01 (0.71)	0.98 (1.09)
Constant	0.15 (0.07)	-0.17 (0.19)	-0.11 (0.22)	-0.05 (0.40)
$N$	379	351	278	198
adj. $R^2$	0.01	0.01	0.01	-0.01
p-values:				
$H_0 : \text{Constant} + I(\tau) = 0$	0.07	0.02	0.85	0.01
$H_0 : \text{Revision} + I(\tau) \times \text{Revision} = 0$	0.27	0.58	0.46	0.74

**Description:** Table shows estimated coefficients from the regression  $y_{i,t+h} - \hat{y}_{i,t+h|\tau} = \alpha_{i,h} + \beta_{i,h}\Delta\hat{y}_{i,t+h|\tau} + \gamma_{i,h}I(\tau) + \lambda_{i,h}I(\tau)\Delta\hat{y}_{i,t+h|\tau} + e_{i,t+h|\tau}$  for Federal Reserve Board staff projections of annualized quarterly real GDP growth by forecast horizon under “cherry picked” assumptions. From our prespecified baseline we instead: estimate unweighted regressions, increase the threshold for dropping revisions to  $\Delta\hat{y}_{i,t+h|\tau}$  to 0.1 p.p., and set  $I(\tau)$  as 28 days until the start of a regularly-scheduled FOMC meeting. Statistical significance asterisks omitted.

**Interpretation:** Under some “cherry picked” assumptions the models show the staff’s GDP forecasts overrevise, which is the opposite conclusion of our prespecified models.

Table 9: “Robustness Check” that Changes the Weights of the “Cherry Picked” Baseline Still Shows Erroneous Results

Forecast Horizon	-1	0	1	2
$I(\tau)$	-0.08 (0.10)	0.37 (0.23)	0.20 (0.30)	-0.29 (0.42)
Revision	-0.33 (0.12)	0.37 (0.50)	-1.24 (0.65)	-1.07 (1.06)
$I(\tau) \times \text{Revision}$	0.26 (0.18)	-0.17 (0.58)	1.10 (0.72)	1.02 (1.09)
Constant	0.16 (0.07)	-0.10 (0.20)	-0.11 (0.22)	-0.05 (0.40)
$N$	379	351	278	198
adj. $R^2$	0.01	0.01	0.01	-0.01
p-values:				
$H_0 : \text{Constant} + I(\tau) = 0$	0.18	0.01	0.65	0.01
$H_0 : \text{Revision} + I(\tau) \times \text{Revision} = 0$	0.62	0.49	0.66	0.86

**Description:** Table shows estimated coefficients from the regression  $y_{i,t+h} - \hat{y}_{i,t+h|\tau} = \alpha_{i,h} + \beta_{i,h}\Delta\hat{y}_{i,t+h|\tau} + \gamma_{i,h}I(\tau) + \lambda_{i,h}I(\tau)\Delta\hat{y}_{i,t+h|\tau} + e_{i,t+h|\tau}$  for Federal Reserve Board staff projections of annualized quarterly real GDP growth by forecast horizon under “cherry picked” assumptions. From our prespecified baseline we instead: estimate weighted regressions by the number of weeks between forecasts (rounded up), increase the threshold for dropping revisions to  $\Delta\hat{y}_{i,t+h|\tau}$  to 0.1 p.p., and set  $I(\tau)$  as 28 days until the start of a regularly-scheduled FOMC meeting. Statistical significance asterisks omitted.

**Interpretation:** Under some “cherry picked” assumptions, including a new assumption for the weights that is different from our preanalysis plan and the “cherry picked” baseline, the models show the staff’s GDP forecasts overrevise, which is the opposite conclusion of our prespecified models.

Table 10: “Robustness Check” that Changes the Minimum Revision Threshold of the “Cherry Picked” Baseline Still Shows Erroneous Results

Forecast Horizon	-1	0	1	2
$I(\tau)$	-0.01 (0.06)	0.23 (0.12)	0.14 (0.16)	0.13 (0.17)
Revision	-0.30 (0.12)	0.20 (0.43)	-1.02 (0.65)	-0.38 (1.00)
$I(\tau) \times \text{Revision}$	0.15 (0.17)	-0.07 (0.49)	0.72 (0.72)	0.17 (1.04)
Constant	0.09 (0.06)	-0.06 (0.10)	-0.38 (0.14)	-0.70 (0.14)
$N$	941	903	723	666
adj. $R^2$	0.00	0.00	0.00	-0.00
p-values:				
$H_0 : \text{Constant} + I(\tau) = 0$	0.01	0.00	0.00	0.00
$H_0 : \text{Revision} + I(\tau) \times \text{Revision} = 0$	0.21	0.58	0.33	0.44

**Description:** Table shows estimated coefficients from the regression  $y_{i,t+h} - \hat{y}_{i,t+h|\tau} = \alpha_{i,h} + \beta_{i,h}\Delta\hat{y}_{i,t+h|\tau} + \gamma_{i,h}I(\tau) + \lambda_{i,h}I(\tau)\Delta\hat{y}_{i,t+h|\tau} + e_{i,t+h|\tau}$  for Federal Reserve Board staff projections of annualized quarterly real GDP growth by forecast horizon under “cherry picked” assumptions. From our prespecified baseline we instead: estimate unweighted regressions, use all revisions in  $\Delta\hat{y}_{i,t+h|\tau}$ , and set  $I(\tau)$  as 28 days until the start of a regularly-scheduled FOMC meeting. Statistical significance asterisks omitted.

**Interpretation:** Under some “cherry picked” assumptions, including a new assumption for the sample restriction on  $\Delta\hat{y}_{i,t+h|\tau}$  that was different from our preanalysis plan and the “cherry picked” baseline, the models show the staff’s GDP forecasts overrevise, which is the opposite conclusion of our prespecified models.

Table 11: “Robustness Check” that Changes the Definition of  $I(\tau)$  of the “Cherry Picked” Baseline Still Shows Erroneous Results

Forecast Horizon	-1	0	1	2
$I(\tau)$	-0.06 (0.09)	0.45 (0.18)	0.20 (0.25)	-0.03 (0.29)
Revision	-0.22 (0.12)	0.17 (0.36)	-1.01 (0.41)	-0.64 (0.53)
$I(\tau) \times \text{Revision}$	0.05 (0.18)	-0.03 (0.45)	0.97 (0.54)	0.59 (0.60)
Constant	0.14 (0.06)	-0.15 (0.14)	-0.16 (0.17)	-0.30 (0.24)
$N$	379	351	278	198
adj. $R^2$	0.01	0.01	0.02	-0.01
p-values:				
$H_0 : \text{Constant} + I(\tau) = 0$	0.14	0.01	0.84	0.04
$H_0 : \text{Revision} + I(\tau) \times \text{Revision} = 0$	0.19	0.59	0.90	0.85

**Description:** Table shows estimated coefficients from the regression  $y_{i,t+h} - \hat{y}_{i,t+h|\tau} = \alpha_{i,h} + \beta_{i,h}\Delta\hat{y}_{i,t+h|\tau} + \gamma_{i,h}I(\tau) + \lambda_{i,h}I(\tau)\Delta\hat{y}_{i,t+h|\tau} + e_{i,t+h|\tau}$  for Federal Reserve Board staff projections of annualized quarterly real GDP growth by forecast horizon under “cherry picked” assumptions. From our prespecified baseline we instead: estimate unweighted regressions, increase the threshold for dropping revisions to  $\Delta\hat{y}_{i,t+h|\tau}$  to 0.1 p.p., and set  $I(\tau)$  as 21 days until the start of a regularly-scheduled FOMC meeting. Statistical significance asterisks omitted.

**Interpretation:** Under some “cherry picked” assumptions, including a new assumption for  $I(\tau)$  that was different from our preanalysis plan and the “cherry picked” baseline, the models show the staff’s GDP forecasts overrevise, which is the opposite conclusion of our prespecified models.

## References

- Arai, Natsuki. 2016. “Evaluating the Efficiency of the FOMC’s New Economic Projections.” *Journal of Money, Credit and Banking* 48(5):1019–1049.
- Aruoba, S. Boragan. 2008. “Data Revisions Are Not Well Behaved.” *Journal of Money, Credit and Banking* 40(2-3):319–340.
- Bauer, Michael D. & Eric T. Swanson. 2020. The Fed’s Response to Economic News Explains the “Fed Information Effect”. Working Paper 27013 National Bureau of Economic Research.
- Berge, Travis J., Andrew C. Chang & Nitish R. Sinha. 2019. “Evaluating the Conditionality of Judgmental Forecasts.” *International Journal of Forecasting* 35(4):1627–1635.
- Bernanke, Ben S. & Kenneth N. Kuttner. 2005. “What Explains the Stock Market’s Reaction to Federal Reserve Policy?” *The Journal of Finance* 60(3):1221–1257.
- Blanco-Perez, Cristina & Abel Brodeur. 2020. “Publication Bias and Editorial Statement on Negative Findings.” *The Economic Journal* 130(629):1226–1247.
- Bloomberg Finance LP. 2017. “Bloomberg Terminals (Open, Anywhere, and Disaster Recovery Licenses).”.
- Board of Governors of the Federal Reserve System. 2000. “Monetary Policy Report, February 17<sup>th</sup> 2000.”.
- Brodeur, Abel, Mathias Lé, Marc Sangnier & Yanos Zylberberg. 2016. “Star Wars: The Empirics Strike Back.” *American Economic Journal: Applied Economics* 8(1):1–32.
- Brodeur, Abel, Nikolai Cook & Anthony Heyes. Forthcoming. “Methods Matter: P-Hacking and Publication Bias in Causal Analysis in Economics.” *American Economic Review* .

- Campbell, Jeffrey R., Charles L. Evans, Jonas D. M. Fisher & Alejandro Justiniano. 2012. “Macroeconomic Effects of Federal Reserve Forward Guidance.” *Brookings Papers on Economic Activity* pp. 1–80.
- Campbell, Sean D. & Steven A. Sharpe. 2009. “Anchoring Bias in Consensus Forecasts and Its Effect on Market Prices.” *Journal of Financial and Quantitative Analysis* 44(2):369–390.
- Casey, Katherine, Rachel Glennerster & Edward Miguel. 2012. “Reshaping Institutions: Evidence on Aid Impacts Using a Preanalysis Plan.” *The Quarterly Journal of Economics* 127(4):1755–1812.
- Chang, Andrew C. & Phillip Li. 2017. “A Preanalysis Plan to Replicate Sixty Economics Research Papers that Worked Half of the Time.” *American Economic Review* 107(5):60–64.
- Chang, Andrew C. & Phillip Li. 2018. “Measurement Error in Macroeconomic Data and Economics Research: Data Revisions, Gross Domestic Product, and Gross Domestic Income.” *Economic Inquiry* 56(3):1846–1869.
- Chang, Andrew C. & Phillip Li. Forthcoming. “Is Economics Research Replicable? Sixty Published Papers From Thirteen Journals Say “Often Not”.” *Critical Finance Review* .
- Chang, Andrew C. & Tyler J. Hanson. 2016. “The Accuracy of Forecasts Prepared for the Federal Open Market Committee.” *Journal of Economics and Business* 83:23–43.
- Chen, Andrew Y. & Tom Zimmermann. 2020. “Publication Bias and the Cross-Section of Stock Returns.” *The Review of Asset Pricing Studies* 10(2):249–289.
- Coibion, Olivier & Yuriy Gorodnichenko. 2012. “What Can Survey Forecasts Tell Us About Information Rigidities?” *Journal of Political Economy* 120(1):116–159.

- Coibion, Olivier & Yuriy Gorodnichenko. 2015. “Is the Phillips Curve Alive and Well after All? Inflation Expectations and the Missing Disinflation.” *American Economic Journal: Macroeconomics* 7(1):197–232.
- Croushore, Dean. 2011. “Frontiers of Real-Time Data Analysis.” *Journal of Economic Literature* 49(1):72–100.
- Croushore, Dean & Simon Van Norden. 2018. “Fiscal Forecasts at the FOMC: Evidence from the Greenbooks.” *Review of Economics and Statistics* 100(5):933–945.
- Croushore, Dean & Simon Van Norden. 2019. “Fiscal Surprises at the FOMC.” *International Journal of Forecasting* 35(4):1583–1595.
- Croushore, Dean & Tom Stark. 2001. “A Real-Time Data Set for Macroeconomists.” *Journal of Econometrics* 105(1):111–130.
- Croushore, Dean & Tom Stark. 2003. “A Real-Time Data Set for Macroeconomists: Does the Data Vintage Matter?” *Review of Economics and Statistics* 85(3):605–617.
- Ericsson, Neil R., Steadman B. Hood, Fred Joutz, Tara M. Sinclair & Herman O. Stekler. 2015. “Time-dependent Bias in the Fed’s Greenbook Forecasts.” *JSM Proceedings, Business and Economics Statistics Section, Alexandria, Virginia* pp. 1568–1582.
- Faust, Jon, John H. Rogers & Jonathan H. Wright. 2005. “News and Noise in G-7 GDP Announcements.” *Journal of Money, Credit and Banking* 37(3):403–419.
- Joutz, Fred & H.O. Stekler. 2000. “An Evaluation of the Predictions of the Federal Reserve.” *International Journal of Forecasting* 16(1):17–38.
- Kahneman, Daniel & Amos Tversky. 1977. *Intuitive Prediction: Biases and Corrective Procedures*. Technical Report PTR 1042-77-6 Defense Advanced Research Projects Agency.
- Koenig, Evan F., Sheila Dolmas & Jeremy Piger. 2003. “The Use and Abuse of Real-Time Data in Economic Forecasting.” *Review of Economics and Statistics* 85(3):618–628.

- Landefeld, J. Steven, Eugene P. Seskin & Barbara M. Fraumeni. 2008. "Taking the Pulse of the Economy: Measuring GDP." *Journal of Economic Perspectives* 22(2):193–216.
- Messina, Jeffrey D., Tara M. Sinclair & Herman Stekler. 2015. "What Can We Learn from Revisions to the Greenbook Forecasts?" *Journal of Macroeconomics* 45:54–62.
- Mincer, Jacob A. & Victor Zarnowitz. 1969. The Evaluation of Economic Forecasts. In *Economic Forecasts and Expectations: Analysis of Forecasting Behavior and Performance*. NBER pp. 3–46.
- Nakamura, Emi & Jón Steinsson. 2018. "High-frequency Identification of Monetary Non-Neutrality: The Information Effect." *The Quarterly Journal of Economics* 133(3):1283–1330.
- Neumark, David. 1999. The Employment Effects of Recent Minimum Wage Increases: Evidence from a Pre-specified Research Design. Working Paper 7171 National Bureau of Economic Research.
- Neumark, David. 2001. "The Employment Effects of Minimum Wages: Evidence from a Prespecified Research Design." *Industrial Relations: A Journal of Economy and Society* 40(1):121–144.
- Reifschneider, David & Peter Tulip. 2019. "Gauging the Uncertainty of the Economic Outlook Using Historical Forecasting Errors: The Federal Reserve's Approach." *International Journal of Forecasting* 35(4):1564–1582.
- Romer, Christina D. & David H. Romer. 2000. "Federal Reserve Information and the Behavior of Interest Rates." *American Economic Review* 90(3):429–547.
- Romer, Christina D. & David H. Romer. 2008. "The FOMC versus the Staff: Where Can Monetary Policymakers Add Value?" *American Economic Review* 98(2):230–35.

- Scotti, Chiara. 2016. “Surprise and Uncertainty Indexes: Real-time Aggregation of Real-Activity Macro-Surprises.” *Journal of Monetary Economics* 82:1–19.
- Sercu, Piet, Martina Vandebroek & Tom Vinaimont. 2008. “Thin-Trading Effects in Beta: Bias v. Estimation Error.” *Journal of Business Finance & Accounting* 35(9-10):1196–1219.
- Tulip, Peter. 2009. “Has the Economy Become More Predictable? Changes in Greenbook Forecast Accuracy.” *Journal of Money, Credit and Banking* 41(6):1217–1231.
- Tversky, Amos & Daniel Kahneman. 1974. “Judgment Under Uncertainty: Heuristics and Biases.” *Science* 185(4157):1124–1131.
- Vilhuber, Lars, James Turitto & Keesler Welch. 2020. “Report by the AEA Data Editor.” *AEA Papers and Proceedings* 110:764–775.
- Vivalt, Eva. 2019. “Specification Searching and Significance Inflation Across Time, Methods and Disciplines.” *Oxford Bulletin of Economics and Statistics* 81(4):797–816.
- White, Halbert. 1980. “A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity.” *Econometrica* 48(4):817–838.
- Woodford, Michael. 2005. Central Bank Communication and Policy Effectiveness. Working Paper 11898 National Bureau of Economic Research.

# A Appendix: High-frequency Forecast Source Data Documentation

This appendix provides a few additional details on the archived documents that we used to create our high-frequency forecast dataset.

The forecasts from the archived documents differ in their precision level, depending on the document type. Briefing tables and charts usually report GDP and inflation at the nearest  $\frac{1}{10}$  percentage point. The eve-of-GDP-release database snapshots usually have the same precision as Greenbooks: GDP at the nearest  $\frac{1}{10}$  percentage point and inflation at the nearest  $\frac{1}{100}$  percentage point. Irregular database backups have forecasts to several decimals, but we rounded these to match the Greenbook’s precision.

The precision of forecasts from briefing texts and forecast update memos varies considerably, but often GDP forecasts from briefing texts and forecast update memos are reported at the nearest  $\frac{1}{4}$  percentage point and inflation forecasts are reported at either the nearest  $\frac{1}{10}$  or  $\frac{1}{4}$  percentage point. These forecasts tend to be paired with a qualifier adjective. For example, consider the briefing text from October 10, 2006 (emphasis ours): “To sum up, the average rate of GDP growth in the second half of the year looks to be about the same as in the September Greenbook, with the third quarter rate —at  $1\frac{1}{4}$  percent —a little slower than we had in the Greenbook, and the fourth quarter pace —at *just over 2 percent* —a little faster.”

In cases where there was a qualifier adjective that indicated “more than”, we applied it to the numerical value in the text and adjusted the forecast in our dataset upward to the nearest  $\frac{1}{10}$  percentage point. For example, “just over 2 percent” is 2.1 percent in our dataset. We recorded adjectives that mean “less than” analogously, by adjusting downward to the nearest  $\frac{1}{10}$  point. When we observed adjectives synonymous with “about,” we did not adjust the forecast.

Occasionally, briefing texts and forecast update memos contain range forecasts. For

example, in the prepared remarks from September 24, 2001 (emphasis ours): “For the fourth quarter, we are likely to project a decline in real GDP of *between one-half and one percent.*” In these cases, we recorded the midpoint of the range as the forecast. We then rounded to match the Greenbook’s precision.

The materials the staff use to brief the Board of Governors —briefing tables and charts that accompany the briefing text —may or may not contain the numerical values of the staff forecasts.<sup>34</sup> But when both the tables or charts and the briefing text report the same forecast, they may also do so at different precision levels. When both the briefing tables and charts and the briefing text contain the same forecast but report at different precision levels, we used the more precise value from the tables and charts.

The eve-of-GDP-release database snapshots have a “previous value” of the staff’s forecast indicated, but the previous value can either be the last Greenbook forecast or the staff’s updated (non-Greenbook) forecast as of the eve-of-GDP-release. There are no metadata that allow us to differentiate between these two possibilities. Furthermore, eve-of-GDP-release database snapshots typically, but not always, report forecasts at the Greenbook’s precision level.

To use eve-of-GDP-release database snapshots we assumed that if the values from the eve-of-GDP-release database snapshot —rounded to the Greenbook’s precision —matched the forecasted values from the last Greenbook, then the eve-of-GDP-release database snapshot reported the Greenbook forecasts. Otherwise we assumed the eve-of-GDP-release database snapshot reflected the staff’s eve-of-GDP-release forecast.

Irregular database backups occur at automatic, time-varying intervals. Because the backups occur at points where the staff may not have vetted the forecasts, sometimes the irregular database backup saves what, in our estimation, is a nonsensical value of the staff’s forecast. We removed these nonsensical values. Though we did not have a hard rule for doing so, typically if the implied revision to the staff’s forecast from an irregular database backup was

---

<sup>34</sup>Briefing tables and charts tend to contain values of new data releases, such as the unemployment rate from the last employment report.

on the order of tens of percentage points for GDP or percentage points for inflation, then we removed those values.

Though the unit of time for our high-frequency dataset is daily, the archived documents that we used to create the dataset are mixed-frequency. Typically we only have one staff forecast for a day. But on occasion we observed more than one staff forecast on a given day. In instances where we have more than one forecast on a given day, we adopted the following priority system, from highest to lowest: (1) Greenbooks, (2) forecast update memos, (3) briefing tables/charts/text, (4) eve-of-GDP-release database snapshots, and (5) automatic database backups.

Greenbooks are the official staff forecast and undergo the most thorough vetting, so we gave them the highest priority. Forecast update memos, briefing tables, charts, and text also contain vetted forecasts because they are direct communications to the Board of Governors, but the staff usually disclaims that the forecasts are still in flux and the reported forecasts could change—only the Greenbooks are the official staff forecast. Eve-of-GDP-release database snapshots also undergo some vetting in preparation for the BEA’s release, though our suspicion is that the forecasts are less vetted than memos or briefing tables/charts/text because the eve-of-GDP-release database snapshots are not official communications to the Board of Governors.

## B Appendix: Outlier Check and Bloomberg Data Tables

Table 12: Removing Outliers Still Suggest Some GDP Forecast Inefficiencies

Forecast Horizon	-1	0	1	2
Revision	-0.11 (0.11)	0.48 (0.29)	-0.18 (0.33)	-0.14 (0.23)
Constant	0.02 (0.04)	0.22 (0.08)	-0.27 (0.13)	-0.60 (0.13)
<i>N</i>	610	574	491	378
adj. <i>R</i> <sup>2</sup>	0.00	0.01	-0.00	-0.00

**Description:** Table shows estimated coefficients from the regression  $y_{i,t+h} - \hat{y}_{i,t+h|\tau} = \alpha_{i,h} + \beta_{i,h} \Delta \hat{y}_{i,t+h|\tau} + e_{i,t+h|\tau}$  for Federal Reserve Board staff projections of annualized quarterly real GDP growth by forecast horizon, removing 1 percent of outliers. We weight by number of days between forecast revisions in  $\Delta \hat{y}_{i,t+h|\tau}$ , with Huber-White (White 1980) standard errors in parentheses. Hypothesis tests are two sided. Statistical significance asterisks omitted.

**Interpretation:** The regressions without outliers still suggest, on average, some forecast inefficiencies. The current-quarter GDP forecasts tend to underrevise, and there is also some evidence of bias.

Table 13: Removing Outliers Still Suggest Limited Inflation Forecast Inefficiencies

Forecast Horizon	-1	0	1
Revision	-0.45 (0.21)	0.23 (0.25)	-1.82 (0.82)
Constant	-0.01 (0.03)	0.00 (0.04)	0.26 (0.10)
$N$	165	204	48
adj. $R^2$	0.10	0.00	0.12

**Description:** Table shows estimated coefficients from the regression  $y_{i,t+h} - \hat{y}_{i,t+h|\tau} = \alpha_{i,h} + \beta_{i,h} \Delta \hat{y}_{i,t+h|\tau} + e_{i,t+h|\tau}$  for Federal Reserve Board staff projections of annualized quarterly core PCE inflation by forecast horizon, removing 1 percent of outliers. We weight by number of days between forecast revisions in  $\Delta \hat{y}_{i,t+h|\tau}$ , with Huber-White (White 1980) standard errors in parentheses. Hypothesis tests are two sided. Statistical significance asterisks omitted.

**Interpretation:** The regressions without outliers still suggest, on average, one-quarter backcasts tend to overrevise but otherwise there is limited evidence of inefficiencies. The one-quarter-ahead regressions have a small sample size, so we do not want to overemphasize their results.

Table 14: Time-Varying Real GDP Efficiency Regressions without Outliers Also Suggest Some Forecast Inefficiencies

Forecast Horizon	-1	0	1	2
$I(\tau)$	-0.02 (0.08)	0.15 (0.15)	0.57 (0.34)	0.56 (0.25)
Revision	-0.16 (0.13)	0.53 (0.37)	-0.10 (0.43)	-0.55 (0.31)
$I(\tau) \times \text{Revision}$	0.17 (0.26)	-0.71 (0.48)	-0.16 (0.65)	0.82 (0.45)
Constant	0.03 (0.05)	0.18 (0.11)	-0.48 (0.14)	-0.87 (0.19)
$N$	610	574	491	378
adj. $R^2$	0.00	0.01	0.01	0.01
p-values:				
$H_0 : \text{Constant} + I(\tau) = 0$	0.88	0.00	0.76	0.05
$H_0 : \text{Revision} + I(\tau) \times \text{Revision} = 0$	0.99	0.56	0.59	0.40

**Description:** Table shows estimated coefficients from the regression  $y_{i,t+h} - \hat{y}_{i,t+h|\tau} = \alpha_{i,h} + \beta_{i,h}\Delta\hat{y}_{i,t+h|\tau} + \gamma_{i,h}I(\tau) + \lambda_{i,h}I(\tau)\Delta\hat{y}_{i,t+h|\tau} + e_{i,t+h|\tau}$  for Federal Reserve Board staff projections of annualized quarterly real GDP growth by forecast horizon, where  $I(\tau)$  is an indicator for a forecast made within 14 calendar days from the start of a regularly-scheduled FOMC meeting, removing 1 percent of outliers. We weight these regressions by number of days between forecast revisions in  $\Delta\hat{y}_{i,t+h|\tau}$ , with Huber-White (White 1980) standard errors in parentheses. Hypothesis tests are two sided. Statistical significance asterisks omitted.

**Interpretation:** Excluding outliers, there is still some evidence that Federal Reserve Board staff GDP forecasts made at least 14 days from the start of a regularly scheduled FOMC meeting underrevise and two-quarter-ahead forecasts overrevise, though the standard errors are somewhat large. There is some stronger evidence of bias.

Table 15: Time-Varying Inflation Efficiency Regressions Without Outliers Also Indicate Limited Forecast Inefficiencies

Forecast Horizon	-1	0	1
$I(\tau)$	-0.00 (0.05)	-0.01 (0.08)	-0.07 (0.18)
Revision	-0.24 (0.39)	0.07 (0.26)	2.32 (1.55)
$I(\tau) \times \text{Revision}$	-0.24 (0.45)	0.41 (0.54)	-4.82 (1.62)
Constant	-0.01 (0.04)	0.01 (0.05)	0.32 (0.15)
$N$	165	204	48
adj. $R^2$	0.09	0.00	0.25
p-values:			
$H_0 : \text{Constant} + I(\tau) = 0$	0.67	0.98	0.02
$H_0 : \text{Revision} + I(\tau) \times \text{Revision} = 0$	0.04	0.31	0.00

**Description:** Table shows estimated coefficients from the regression  $y_{i,t+h} - \hat{y}_{i,t+h|\tau} = \alpha_{i,h} + \beta_{i,h} \Delta \hat{y}_{i,t+h|\tau} + \gamma_{i,h} I(\tau) + \lambda_{i,h} I(\tau) \Delta \hat{y}_{i,t+h|\tau} + e_{i,t+h|\tau}$  for Federal Reserve Board staff projections of annualized quarterly core PCE inflation by forecast horizon, where  $I(\tau)$  is an indicator for a forecast made within 14 days from the start of a regularly-scheduled FOMC meeting, removing 1 percent of outliers. We weight these regressions by number of days between forecast revisions in  $\Delta \hat{y}_{i,t+h|\tau}$ , with Huber-White (White 1980) standard errors in parentheses. Hypothesis tests are two sided. Statistical significance asterisks omitted.

**Interpretation:** Excluding outliers, there is still some evidence that Federal Reserve Board staff inflation backcasts made within 14 days from the start of a regularly scheduled FOMC meeting overrevised. Otherwise, there is not much evidence of inefficient inflation forecasts. The sample size of one-quarter ahead inflation regressions is small, so we do not place too much weight on these results.

Table 16: Excluding Outliers Better-than-Expected Macroeconomic News Still Predicts Staff GDP Forecast Errors

Forecast Horizon	-1	0	1	2
$I(\tau)$	-0.01 (0.08)	0.23 (0.15)	0.47 (0.33)	0.27 (0.26)
Revision	-0.16 (0.13)	0.53 (0.37)	-0.36 (0.41)	-0.75 (0.30)
$I(\tau) \times$ Revision	0.13 (0.26)	-0.52 (0.52)	-0.49 (0.65)	1.09 (0.45)
$news_\tau$	-0.02 (0.01)	-0.03 (0.01)	-0.06 (0.03)	-0.06 (0.03)
$I(\tau) \times news_\tau$	0.01 (0.01)	-0.03 (0.02)	-0.02 (0.05)	0.07 (0.04)
Constant	0.04 (0.05)	0.16 (0.10)	-0.40 (0.12)	-0.70 (0.18)
$N$	610	574	491	378
adj. $R^2$	0.02	0.04	0.07	0.03
p-values:				
$H_0 : \text{Constant} + I(\tau) = 0$	0.64	0.00	0.82	0.03
$H_0 : \text{Revision} + I(\tau) \times \text{Revision} = 0$	0.89	0.97	0.09	0.31
$H_0 : news_\tau + I(\tau) \times news_\tau = 0$	0.13	0.00	0.04	0.64

**Description:** Table shows estimated coefficients from the regression  $y_{i,t+h} - \hat{y}_{i,t+h|\tau} = \alpha_{i,h} + \beta_{i,h} \Delta \hat{y}_{i,t+h|\tau} + \gamma_{i,h} I(\tau) + \lambda_{i,h} I(\tau) \Delta \hat{y}_{i,t+h|\tau} + \eta_{i,h} news_\tau + \theta_{i,h} I(\tau) news_\tau + e_{i,t+h|\tau}$  for Federal Reserve Board staff projections of annualized quarterly real GDP growth by forecast horizon, where  $I(\tau)$  is an indicator for a forecast made within 14 days from the start of a regularly-scheduled FOMC meeting, and  $news_\tau$  is our measure of the market's reaction to macroeconomic news using data from Bloomberg Finance LP (2017). We weight these regressions by number of days between forecast revisions in  $\Delta \hat{y}_{i,t+h|\tau}$ , with Huber-White (White 1980) standard errors in parentheses. Hypothesis tests are two sided. Statistical significance asterisks omitted.

**Interpretation:** Excluding outliers, we still find evidence that the market's reaction to macroeconomic news predicts Federal Reserve Board staff real GDP forecast errors, suggesting that the staff does not use information in asset price changes efficiently to inform its GDP forecasts. When economic news is better than expected, the staff forecasts of GDP are more accurate.

Table 17: Excluding Outliers Better-than-Expected Macroeconomic News Still Does Not Predict Staff Inflation Forecast Errors

Forecast Horizon	-1	0	1
$I(\tau)$	0.05 (0.06)	-0.00 (0.08)	-0.19 (0.17)
Revision	-0.17 (0.40)	0.09 (0.25)	1.43 (1.51)
$I(\tau) \times \text{Revision}$	-0.34 (0.47)	0.34 (0.54)	-3.90 (1.62)
$news_\tau$	-0.00 (0.01)	-0.00 (0.01)	-0.04 (0.01)
$I(\tau) \times news_\tau$	-0.01 (0.01)	-0.01 (0.01)	0.03 (0.01)
Constant	-0.05 (0.05)	0.02 (0.05)	0.50 (0.13)
$N$	165	204	48
adj. $R^2$	0.09	0.01	0.40
p-values:			
$H_0 : \text{Constant} + I(\tau) = 0$	0.95	0.80	0.01
$H_0 : \text{Revision} + I(\tau) \times \text{Revision} = 0$	0.03	0.38	0.00
$H_0 : news_\tau + I(\tau) \times news_\tau = 0$	0.27	0.20	0.00

**Description:** Table shows estimated coefficients from the regression  $y_{i,t+h} - \hat{y}_{i,t+h|\tau} = \alpha_{i,h} + \beta_{i,h} \Delta \hat{y}_{i,t+h|\tau} + \gamma_{i,h} I(\tau) + \lambda_{i,h} I(\tau) \Delta \hat{y}_{i,t+h|\tau} + \eta_{i,h} news_\tau + \theta_{i,h} I(\tau) news_\tau + e_{i,t+h|\tau}$  for Federal Reserve Board staff projections of annualized quarterly core PCE inflation by forecast horizon, where  $I(\tau)$  is an indicator for a forecast made within 14 days from the start of a regularly-scheduled FOMC meeting, and  $news_\tau$  is our measure of the market's reaction to macroeconomic news using data from Bloomberg Finance LP (2017). We weight these regressions by number of days between forecast revisions in  $\Delta \hat{y}_{i,t+h|\tau}$ , with Huber-White (White 1980) standard errors in parentheses. Hypothesis tests are two sided. Statistical significance asterisks omitted.

**Interpretation** Excluding outliers, we still do not find evidence that the market's reaction to macroeconomic news predicts Federal Reserve Board staff inflation forecast errors.

Table 18: Bloomberg Series in News Index,  $news_{\tau}$  (Table 1 of 2)

Bloomberg Mnemonic	Data Description
ADP_CHNG_INDEX	ADP National Employment Report, SA, Private Nonfarm Level Change
ADV_GDP_CQOQ_INDEX	US GDP First Release, Chained, QoQ, SAAR
ADV_GDP_PIQQ_INDEX	US GDP Price Index First Release, QoQ, SAAR
AHE_MOM%_INDEX	US Average Hourly Earnings All Employees, Total Private Monthly Percentage Change
CHPMINDEX_INDEX	MNI Chicago Business Barometer SA
CICRTOT_INDEX	Federal Reserve G19 Consumer Credit Total Net Change SA
CNSTTMOM_INDEX	Census Bureau US Construction Spending MoM SA
CONCCONF_INDEX	Conference Board Consumer Confidence SA
CONSENT_INDEX	University of Michigan, Survey of Consumer Confidence Sentiment (Final)
COSTNFR%_INDEX	US Unit Labor Costs Nonfarm Business Sector QoQ % SAAR
CPI_CHNG_INDEX	US CPI Urban Consumers MoM SA
CPI_XYOY_INDEX	US CPI Urban Consumers Less Food & Energy YoY NSA
CPI_YOY_INDEX	US CPI Urban Consumers YoY NSA
CPTICHNG_INDEX	US Capacity Utilization % of Total Capacity SA
CPUPXCHG_INDEX	US CPI Urban Consumers Less Food & Energy MoM SA
DGNOCHNG_INDEX	US Durable Goods New Orders Industries MoM SA
ECL_SA%_INDEX	Bureau of Labor Statistics, Employment Cost Civilian Workers QoQ SA
EMPRGBCL_INDEX	Empire State Manufacturing Survey, General Business Conditions SA
ETSLMOM_INDEX	US Existing Homes Sales MoM SA
FDDSSD_INDEX	US Treasury Federal Budget Debt Summary, Deficit Or Surplus NSA
FDTR_INDEX	Federal Funds Target Rate Upper Bound, p.p.
FRNTTOTL_INDEX	US Foreign Net Transactions
GDPCPCEC_INDEX	US GDP Personal Consumption Core Price Index, QoQ % SAAR
GDPCTOT%_INDEX	GDP US Personal Consumption Chained, % Change from Previous Period SAAR
GDP_CQOQ_INDEX	US GDP Third Release, Chained, QoQ, SAAR
GDP_PIQQ_INDEX	US GDP Price Index Third Release, QoQ, SAAR
HPIMMOM%_INDEX	FHFA US House Price Index Purchase Only MoM% SA
IMP1CHNG_INDEX	US Import Price Index by End Use All MoM NSA
INJCJC_INDEX	US Initial Jobless Claims SA
INJCSP_INDEX	US Continuing Jobless Claims SA
IP_CHNG_INDEX	Industrial Production, Change from Previous Period, SA
LELCHNG_INDEX	Conference Board US Leading Index MoM

**Notes:** MoM = month over month, QoQ = quarter over quarter, SA = seasonally adjusted, SAAR = seasonally adjusted at an annual rate, YoY = year over year. Source: Bloomberg Finance LP (2017).

Table 19: Bloomberg Series in News Index,  $news_{\tau}$  (Table 2 of 2)

Bloomberg Mnemonic	Data Description
MWINCHNG_INDEX	Merchant Wholesalers Inventories Total Monthly % Change
NAPMNMI_INDEX	ISM Non-Manufacturing NMI Composite
NAPMPMI_INDEX	ISM Manufacturing PMI SA
NFP_PCH_INDEX	US Employees on Nonfarm Payrolls, Total Private MoM Net Change SA
NFP_TCH_INDEX	US Employees on Nonfarm Payrolls, Total MoM Net Change SA
NHSLTOT_INDEX	US New One Family Houses Sold Annual Total SAAR
NHSPSTOT_INDEX	Housing Starts, SAAR
OUTFGAF_INDEX	Philadelphia Fed Business Outlook Survey, Diffusion Index General Conditions
PCE_CMOM_INDEX	US Personal Consumption Expenditures, Core Price Index MoM SA
PCE_CRCH_INDEX	US Personal Consumption Expenditures, Nominal Dollars MoM SA
PCE_CYOY_INDEX	US Personal Consumption Expenditures, Core Price Index YoY SA
PCE_DEFY_INDEX	US Personal Consumption Expenditures, Chain Type Price Index YoY SA
PITLCHNG_INDEX	US Personal Income MoM SA
PPLCHNG_INDEX	US PPI Finished Goods SA MoM %
PRE_CONSSENT_INDEX	University of Michigan, Survey of Consumer Confidence Sentiment (Preliminary)
PRODNFR%_INDEX	US Output Per Hour Nonfarm Business Sector QoQ SA
PXFECHNG_INDEX	US PPI Finished Goods Less Foods & Energy SA MoM%
RSTAMOM_INDEX	Adjusted Retail & Food Services Sales, SA Total Monthly % Change
RSTAXMOM_INDEX	Adjusted Retail Sales Less Autos SA Monthly % Change
SAARDTOT_INDEX	US Auto Sales Domestic Vehicles Annualized SA
SBOITOTL_INDEX	NFIB Small Business Optimism Index
SEC_GDP_CQOQ_INDEX	US GDP Second Release, Chained, QoQ, SAAR
SEC_GDP_PIQQ_INDEX	US GDP Price Index Second Release, QoQ, SAAR
SPCS20Y%_INDEX	S&P/Case-Shiller Composite-20 City Home Price Index YoY
TMNOCHNG_INDEX	US Manufacturers New Orders Total MoM SA
USCABAL_INDEX	US Nominal Account Balance In Billions of USD
USMMMCH_INDEX	US Employees on Nonfarm Payrolls, Manufacturing Industry Monthly Net Change SA
USPHTMOM_INDEX	US Pending Home Sales Index MoM SA
USTBTOT_INDEX	US Trade Balance Of Payments SA
USURTOT_INDEX	U-3 US Unemployment Rate SA

**Notes:** MoM = month over month, QoQ = quarter over quarter, SA = seasonally adjusted, SAAR = seasonally adjusted at an annual rate, YoY = year over year. Source: Bloomberg Finance LP (2017).