NBER WORKING PAPER SERIES

BELIEF DISTORTIONS AND MACROECONOMIC FLUCTUATIONS

Francesco Bianchi
Sydney C. Ludvigson
Sai Ma

Working Paper 27406
http://www.nber.org/papers/w27406

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2020

We thank Michael Boutros, Justin Shugarman, and Yueteng Zhu for their excellent research assistance. The views expressed are those of the authors and do not necessarily reflect those of the Federal Reserve Board, the Federal Reserve System, or the National Bureau of Economic Research.

Belief Distortions and Macroeconomic Fluctuations
Francesco Bianchi, Sydney C. Ludvigson, and Sai Ma
NBER Working Paper No. 27406
June 2020
JEL No. E03,E17,E7

ABSTRACT

This paper combines a data rich environment with a machine learning algorithm to provide estimates of time-varying systematic expectational errors ("belief distortions") about the macroeconomy embedded in survey responses. We find that such distortions are large on average even for professional forecasters, with all respondent-types over-weighting their own forecast relative to other information. Forecasts of inflation and GDP growth oscillate between optimism and pessimism by quantitatively large amounts. To investigate the dynamic relation of belief distortions with the macroeconomy, we construct indexes of aggregate (across surveys and respondents) expectational biases in survey forecasts. Over-optimism is associated with an increase in aggregate economic activity. Our estimates provide a benchmark to evaluate theories for which information capacity constraints, extrapolation, sentiments, ambiguity aversion, and other departures from full information rational expectations play a role in business cycles.

Francesco Bianchi
Social Sciences Building, 201B
Department of Economics
Duke University
Box 90097
Durham, NC 27708-0097
and CEPR
and also NBER
francesco.bianchi@duke.edu

Sydney C. Ludvigson
Department of Economics
New York University
19 W. 4th Street, 6th Floor
New York, NY 10002
and NBER
sydney.ludvigson@nyu.edu

Sai Ma
Federal Reserve Board,
C Ave & 20th Street NW
Washington, DC 20551
sai.ma@frb.gov

1 Introduction

How important are belief distortions in economic decision making and what role do they play in macroeconomic fluctuations? Large theoretical literatures have emerged to argue that systematic expectational errors embedded in beliefs can have important dynamic effects on the economy. But much less is known about the empirical relation of any such distortions to macroeconomic activity.

To formalize our notion of "belief distortion," let us define it in general terms as *an expectational error generated by the systematic mis-weighting of available information demonstrably pertinent to the accuracy of the belief*. This definition nests those that consider errors generated by merely omitting relevant information to include any instance where information is suboptimally given too much or too little weight. In the theoretical literatures on distorted beliefs, economic agents make systematic expectational errors for a variety of reasons. These reasons include the presence of information frictions (e.g., see the literature overview in Coibion and Gorodnichenko (2015)), the use of simple extrapolative rules (e.g., De Long, Shleifer, Summers, and Waldmann (1990); Barberis, Shleifer, and Vishny (1998); Barberis, Greenwood, Jin, and Shleifer (2015)), the intentional adoption of conservatively pessimistic beliefs (e.g., the overviews in Hansen and Sargent (2008) and Epstein and Schneider (2010)), the overweighting of personal experiences (e.g., Malmendier and Nagel (2011); Malmendier and Nagel (2015)), the over-reaction to incoming news (e.g., Bordalo, Gennaioli, and Shleifer (2018); Gennaioli and Shleifer (2018)), or the presence of priors with skewness (e.g., Afrouzi, Veldkamp, et al. (2019)), among others.

A challenge in empirically assessing the role of belief distortions, and their relation to macroeconomic activity, is that no objective measure of belief distortions exists. So far, empirical work has largely proceeded by investigating whether the forecast errors made by survey respondents deviate from some benchmark model of nondistorted beliefs. But because the existing studies differ widely according to their empirical design, there is no generally accepted metric of belief distortions that applies across a range of cases. Studies differ according to the specific surveys that are investigated, the segment of the population that is surveyed, the topic of the survey questions, the time period to which the survey questions pertain, and the empirical methodology used to identify systematic errors in expectations. Perhaps most important, given the wide-ranging theoretical literatures cited above and the vast amount of information that could be considered ex-ante known and pertinent to economic decision-making, it is not obvious what benchmark model of beliefs should be applied to measure any distortion in survey responses.

This paper provides new measures of systematic expectational errors in survey responses and relates them to macroeconomic activity. Our objective is to construct and study a comprehensive, methodologically consistent econometric measure of belief distortions in macroeconomic expectations by looking across a range of surveys, a range of agent types, and a range of questions about future economic outcomes. Returning to our definition of belief distortions above, it is clear that such a measurement requires three key ingredients.

First, we require direct evidence on what economic decision-makers actually believe. For this we obtain data from several different surveys, different survey questions, and broad cross-sections of survey respondents with different beliefs. Second, we must cope with the theoretically vast quantity of available information that is possibly pertinent to belief accuracy, as well as account for other bona fide features of real time decision making, such as the out-of-sample nature of foreword-looking judgements. Failure to take into account either the data-rich environment in which survey respondents operate or the out-of-sample nature of their forecasts can lead to erroneous conclusions about belief distortions and their relation to the macroeconomy. For this, we use a machine learning algorithm to process hundreds of pieces of information that would have been available to survey respondents in real time at daily, quarterly, and monthly sampling intervals. The third and final ingredient is the availability of observations on both survey responses and objective economic information over a sufficiently long time span. This is required both to reduce sampling noise, as is necessary to distinguish bad luck in a random environment from a systematic mis-weighting of information, as well as to statistically infer the role of any belief distortions in dynamic macroeconomic fluctuations.

With these ingredients in hand, we ask whether cross-sections of survey respondents with different beliefs systematically mis-weight pertinent economic information by comparing the survey forecasts to a benchmark machine learning forecast. To reduce sampling uncertainty and distinguish ex post bad luck from systematic expectational error, we average the squared errors produced by the machine learning algorithm and the survey forecasts over an extended evaluation sample. If the machine benchmark produces more reliable forecasts on average over an extended sample, we take that as evidence of systematic expectational error. Otherwise we conclude there is no evidence of systematic error.

Machine learning is itself a model of belief formation. We argue that it provides an appropriate benchmark for quantifying biases in survey responses, for several reasons. First, optimized approaches to real world decision and prediction problems almost always require the efficient processing of large amounts of information. This clearly applies to professional forecasters who are presumably among the most informed agents in the economy, but also to other agent-types, including investors, firms, governments, and sometimes even households. In all cases, a benchmark based on a small amount of arbitrarily chosen information could fail to reveal systematic expectational errors or, conversely, lead to spurious evidence of systematic error. Machine algorithms are advantageous in this regard because they are explicitly designed to cope with large amounts of information. Second, to the extent that respondents with heterogenous beliefs differ in their degree of distortion, a suitable machine benchmark can quantify this heterogeneity and the type of information that is mis-weighted over time. Agents with some beliefs may be quite close to the machine efficient benchmark, while others may be quite far from it. Third, the machine can easily be coded to adapt to new information as it becomes available and make out-of-sample forecasts on this basis. Thus, the approach does not run the risk of erroneously indicating that forecaster performance is suboptimal merely because of the existence of structural breaks and/or the arrival of new information that even the most efficient information processing algorithm could have learned about only slowly over time.

As an illustration of the importance of minding these features of real time decision-making, we show that ignoring them can have important effects on the conclusions one would draw. For example, in contrast to well established findings from in-sample regressions, out-of-sample estimation provides little evidence that lagged ex-ante revisions in survey forecasts have predictive power for average survey forecast errors. Moreover, some information found elsewhere to be important for prediction out-of-sample while in a comparatively low-dimensional setting is found to be unimportant in our high-dimensional, data-rich setting.

Our main findings may be summarized as follows. First, when compared to a range of surveys and respondent-types with heterogeneous beliefs, the machine model produces lower mean squared forecast errors of inflation and GDP growth over an extended evaluation sample, sometimes by large margins. The machine model does so by altering the relative weight placed on real-time information versus the survey forecast, while adapting dynamically to changing information as it moves through a forecast evaluation sample. This evidence includes large distortions even in the consensus estimates of professional forecasters. A key finding is that survey respondents of all types place too much weight on their own forecast relative to other information, and are in that sense overconfident.

Second, we find that biases in inflation expectations for the median respondent of all types are on average too high over our evaluation sample, a direction we shall refer to as "pessimistic."[1] By contrast, biases in expectations of economic growth are "optimistic" on average–i.e., too high–for the median respondent among professional forecasters and corporate executives, while they are very slightly pessimistic for households. But these averages mask large variation over time in the median respondent's bias, as well across respondents at any given point in time.

Third, although the machine's dynamic regularization algorithm often results in a sparse specification, this is not so in every time period. And even in time periods where sparse specifications are chosen, the type of information utilized by the specification changes from period to period. These findings underscore the value of a dynamic large-scale learning algorithm for achieving the machine-optimal forecast, even if much of the information is associated with a

[1] An alternative interpretation is that higher expected inflation represents an optimistic view of the world in certain episodes, such as recessions. However, such an interpretation seems to be at odds with surveys of inflation attitudes. See the discussion in Bhandari, Borovicka, and Ho (2019).

coefficient that is shrunk all the way to zero most of the time.

With this evidence in hand, we form a measure of the dynamics of systematic expectational errors by tracking the difference between the survey forecast and the machine forecast over time. Survey forecasts of both inflation and GDP growth fluctuate between optimism and pessimism. For GDP growth, we find extended periods of optimism that are especially prevalent for the median forecast among professional forecasters. For example, from 2010-2018, median professional forecasts of economic growth are biased upward by 0.83% at an annual rate, or 37% of actual GDP growth during this period. For inflation, the median forecast across all respondent types also exhibits extended periods of optimism, despite being pessimistic on average over the full sample. For example, median inflation forecasts exhibit a downward bias from 2011 to 2014 that ranges across respondent-types from –0.34% to -1.03% at an annual rate, or -19% to -47% of actual inflation during this period.

To what extent are time-varying belief distortions correlated with macroeconomic fluctuations? To investigate their dynamic relation with the macroeconomy, we construct indexes of aggregate (across surveys and respondents) inflation and GDP growth biases, placing these indexes in a vector autoregression with macroeconomic data. We find that a positive innovation to the inflation bias index (indicating an increase in pessimism) operates much like a cost-push shock, driving up the real wage and driving down real investment, real GDP, and the price level. By contrast, a positive innovation to the GDP growth bias index (indicating an increase in optimism) has the opposite effect and leads to a sizable and more protracted *in*crease in real activity, the price level, and also the stock market, while the real wage declines. Importantly, these results are specific to innovations in the systematic expectational errors survey respondents make, and not to their expectations *per se*. Indeed, positive innovations to an index of GDP growth *expectations* have very different effects and are not associated with a boom in economic activity or the stock market.

The rest of this paper is organized as follows. Section 2 reviews related literature not discussed above. Section 3 describes our econometric and machine learning framework. Section 4 describes the results and Section 5 concludes. A large amount of additional material on our data construction, estimation, and additional robustness checks have been placed in an Appendix for online publication.

2 Related Empirical Literature

nice lit review

Our estimates provide a benchmark to evaluate theories for which information capacity constraints, extrapolation, sentiments, ambiguity aversion, and other departures from full information rational expectations play a role in business cycles. Apart from the large literatures mentioned above, a related literature on "sentiments" postulates that communication frictions may prevent agents from reaching similar expectations about economic activity, in which case aggregate expectations would exhibit belief distortions as defined above (e.g., Angeletos and La'O (2013); Angeletos, Collard, and Dellas (2018); Milani (2011); Milani (2017)). Similarly, micro-founded models featuring confidence shocks include those with ambiguity averse agents who are deliberately pessimistic on average (e.g., Ilut and Schneider (2015), Bianchi, Ilut, and Schneider (2017), Ilut and Saijo (2020), Bhandari, Borovicka, and Ho (2019)). There remains a question of whether ambiguity aversion would show in survey responses. If not, such models need some other mechanism to explain the systematic expectational errors documented here and elsewhere. The above theories provide a mechanism through which a relatively unbiased machine benchmark operating in a data rich environment would provide forecasts that deviate from those made by humans and possibly be more accurate.

The systematic expectational errors we seek to measure recall a literature in economic psychology that studies how basic properties of cognition may give rise to human biases in expectation formation. Khaw, Stevens, and Woodford (2017) conduct a laboratory experiment to shed light on cognitive limitations that influence how decision makers respond to changes in their economic environment. Woodford (2013) argues that the degree of accuracy with which individuals perceive objective reality can have effects on how expectations are formed.

In terms of our macroeconomic findings, we touch on a large literature that documents the existence of belief distortions, and relates them to economic activity. These papers find evidence of departures from rational expectations in predicting inflation and other macro variables (Coibion and Gorodnichenko 2012, 2015; Fuhrer 2017), the aggregate stock market (Bacchetta, Mertens, and van Wincoop 2009, Amromin and Sharpe 2014, Greenwood and Shleifer 2014, Adam, Marcet, and Buetel 2017), the cross section of stock returns (Bordalo, Gennaioli, La Porta and Shleifer 2017), credit spreads (Greenwood and Hanson 2013, Bordalo, Gennaioli, and Shleifer 2018), and corporate earnings (DeBondt and Thaler 1990, Ben-David et. al. 2013, Gennaioli, Ma, and Shleifer 2015, Bouchaud, Kruger, Landier, and Thesmar 2017). None of these studies take into account the data-rich environment or the out-of-sample nature of decision-making, nor do they investigate the effects of distorted beliefs in a macroeconomic vector autoregression, gaps our study is designed to fill.

Moreover, our findings differ in notable ways from some in the extant literature. For example, following Coibion and Gorodnichenko (2015), we ask whether ex ante revisions in the average forecast reduce average ex post forecast errors, as would be indicative of models with information rigidities. But we find no evidence that they do. Instead, the coefficients on forecast revisions are typically shrunk to zero by the machine algorithm in favor of placing greater absolute weight on other pieces of information. And unlike most models with ambiguity averse agents, we do not find that individuals are necessarily pessimistic on average over longer periods of time. Instead, survey respondents fluctuate between periods of optimism and pessimism, with the results varying by economic indicator and respondent type.

Finally, we are aware of relatively little work that has used machine learning as a benchmark model of belief formation. An important exception is Martin and Nagel (2019) who use it to study models of expected stock returns in the cross-section. Although their context is very different from ours, they find, as we do, that ignoring the interplay between a data-rich environment and genuine ex ante decision making can lead to erroneous conclusions about whether beliefs are rational and information is efficiently weighted.

3 Econometric and Machine Learning Framework

We now turn to a description of our econometric and machine learning framework. Both aspects require a sufficiently long time series of observations, including those on survey responses. Because the panel elements of our survey data are too limited to do the analysis on a respondent-level basis, we work instead with the surveys' repeated cross-sections to form a machine benchmark for respondents in different *percentiles* of the survey forecast distributions over time.

Let $y_{j,t+h}$ generically denote a economic time series indexed by j whose value in period $h \geq 1$ a survey forecaster is asked to predict at time t. Let $\mathbb{F}_t^{(i)}$ generically denote a survey forecast made at time t and let superscript (i) denote either the mean belief, in which case "$i = \mu$", or the respondent located at the ith percentile of the survey forecast distribution, i.e., "$i = 65$" refers to the belief of the respondent at the 65th percentile. Thus $\mathbb{F}_t^{(65)}[y_{j,t+h}]$ denotes the survey expectation of $y_{j,t+h}$ that is formed at time t by the respondent at the 65th percentile of the survey distribution.

In order to identify possible distortions in beliefs, it is imperative that the benchmark model of belief formation be as rich as possible, so that our measure of distortion does not miss pertinent information or pertain only to a small number of arbitrarily chosen information variables. This is especially important in our context since relevant information not considered by the benchmark can lead to spurious estimates of systematic expectational errors.

To address this problem we take a two pronged approach that combines diffusion index estimation with machine learning. The diffusion index estimation component is a preliminary dimension-reduction step wherein a relatively small number of dynamic factors are estimated from hundreds of economic time-series. The approach enables the use of a possibly vast set of economic variables that is more likely to span the information sets available in real time to economic decision makers. Nonlinearities are captured by including polynomial functions of estimated dynamic factors, or by forming additional factors from polynomials of the raw data. The second step in our analysis is to combine diffusion index forecasting with a machine algorithm of regularized estimation to optimally trades off downweighting information with reduced parameter estimation error. Diffusion index forecasting is increasingly used in data rich environments. Thus we touch only briefly on this step and focus instead on the machine learning benchmark, leaving details about estimation of factors to the Online Appendix.

3.1 Machine Learning Benchmark

Let $x_t^C = (x_{1t}^C, \ldots, x_{Nt}^C)'$ generically denote a dataset of economic information in some category C that is available for real-time analysis. It is assumed that x_t^C has been suitably transformed (such as by taking logs and differencing) so as to render the series stationary. We assume that x_{it}^C has an approximate factor structure taking the form

$$x_{it}^C = \Lambda_i^{C\prime} \mathbf{G}_t^C + e_{it}^X,$$

where $\mathbf{G}_t^C$ is an $r_G \times 1$ vector of latent common factors ("diffusion indexes"), $\mathbf{\Lambda}_i^C$ is a corresponding $r_C \times 1$ vector of latent factor loadings, and e_{it}^X is a vector of idiosyncratic errors.[2] The number of factors r_G is typically significantly smaller than the number of series, N, which facilitates the use of very large datasets. Additional factors are formed to account for nonlinearities by taking polynomial functions of factors $\mathbf{G}_t^C$, as well as by forming factors from polynomials of the raw data.

Collect all factors from different datasets of category C, as well as nonlinear components (polynomials of factors and factors formed from polynomials of raw data) into a single r_G dimensional vector $\mathbf{G}_t$. Let $\hat{\mathbf{G}}_t$ denote consistent estimates of a rotation of $\mathbf{G}_t$ and let the r_W dimensional vector $\mathbf{W}_t$ contain additional non-factor information that will be specified below. Finally, let $\mathbf{Z}_{jt} \equiv \left(y_{j,t}, \hat{\mathbf{G}}_t', \mathbf{W}_{jt}'\right)'$ be a $r = 1 + r_G + r_W$ vector which collects the data at time t and let $\mathcal{Z}_{jt} \equiv \left(y_{j,t}, ..., y_{j,t-p_y}, \hat{\mathbf{G}}_t', ..., \hat{\mathbf{G}}_{t-p_G}', \mathbf{W}_{jt}', ..., \mathbf{W}_{jt-p_W}'\right)'$ be a vector of contemporaneous and lagged values of $\mathbf{Z}_{jt}$, where p_y, p_G, p_W denote the total number of lags of $y_{j,t}$, $\hat{\mathbf{G}}_t'$, $\mathbf{W}_{jt}'$, respectively.

With these data, consider the following machine forecasting model for $y_{j,t+h}$:

$$y_{j,t+h} = \alpha_j^{(i)} + \beta_{j\mathbb{F}}^{(i)} \mathbb{F}_t^{(i)} \left[y_{j,t+h}\right] + \underbrace{\mathbf{B}_{j\mathcal{Z}}^{(i)\prime}}_{1\times K} \mathcal{Z}_{jt} + \epsilon_{jt+h}, \qquad h \geq 1 \tag{1}$$

where $K = r + p_y + p_G \cdot r_G + p_W \cdot r_W$ is the number of right-hand-side variables other than $\mathbb{F}_t^{(i)}$, and $\alpha_j^{(i)}$ is an intercept term. We define the *machine efficient benchmark* as a set of parameter restrictions that would imply the survey forecaster in the ith percentile processes all available information at time t as efficiently as the machine. This benchmark corresponds to the following restrictions:

$$\beta_{j\mathbb{F}}^{(i)} = 1; \ \mathbf{B}_{j\mathcal{Z}}^{(i)} = \mathbf{0}; \ \alpha_j^{(i)} = 0. \tag{2}$$

[2] In an approximate dynamic factor structure, the idiosyncratic errors e_{it}^X are permitted to have a limited amount of cross-sectional correlation.

If the machine-efficient forecasts are more accurate over an extended time period than the survey respondent's forecasts, we take that as evidence of systematic expectational error in the survey forecast. Such errors would then be revealed by deviations from the above benchmark, generated by a mis-weighting of information contained in $\mathcal{Z}_{jt}$ or "1" (i.e., $\mathbf{B}_{j\mathcal{Z}}^{(i)} \neq \mathbf{0}$ or $\alpha_j^{(i)} \neq 0$) and/or the survey respondent's own forecast, $\mathbb{F}_t^{(i)}[y_{j,t+h}]$ (i.e., $\beta_{j\mathbb{F}}^{(i)} \neq 1$).

Two points about this benchmark bear noting. First, the machine learning model of belief formation is a *percentile-specific* benchmark that adopts the perspective of a forecaster at the ith percentile in period t. The machine is given any information that the survey forecaster could have observed at time t, including her own forecast $\mathbb{F}_t^{(i)}[y_{j,t+h}]$, as well as all objective economic information contained in $\mathcal{Z}_t$, which may include lagged values of her own or other respondent's forecasts $\mathbb{F}_{t-1}^{(s\neq i)}[y_{j,t+h}]$. Note that the benchmark assumes the respondent in the ith percentile knows her own "type," so that she has a sense of where in the time t forecast distribution her response is located. Second, when $\beta_{j\mathbb{F}}^{(i)}$ differs from unity, the above benchmark implies that the belief $\mathbb{F}_t^{(i)}[y_{j,t+h}]$ could have been improved by reweighting the respondent's own forecast against other information contained in $\mathcal{Z}_t$.

There are potentially thousands of pieces of information that could be considered in the machine learning model (1). We use dynamic factors to substantially reduce the dimensionality of the estimation problem. But even with this first-stage dimension-reduction step, the number of possible predictors on the right-hand-side of (1) can still be quite large, possibly exceeding the number of real-time observations available to estimate the relation of $y_{j,t+h}$ with information variables, especially when rolling subsamples are used as part of a dynamic learning algorithm. In this setting, high degrees of parameter estimation error and over-fitting are likely even with the aid of dynamic factors. Our next step is to therefore use machine learning along with data driven regularization to address the high-dimensional benchmark learning problem.

3.2 Machine Learning Algorithm

To simplify notation, collect all the independent variables and coefficients on the right-hand-side of (1) into a single matrix and vector and write the machine predictive model as:

$$y_{j,t+h} = \mathcal{X}_t' \boldsymbol{\beta}_j^{(i)} + \epsilon_{jt+h} \tag{3}$$

where $\mathcal{X}_t = \left(1, \mathbb{F}_t^{(i)}[y_{j,t+h}], \mathcal{Z}_{jt}\right)'$ and $\boldsymbol{\beta}_j^{(i)} \equiv \left(\alpha_j^{(i)}, \beta_{j\mathbb{F}}^{(i)}, \left(\mathbf{B}_{j\mathcal{Z}}^{(i)}\right)\right)'$. Let $\mathbf{X}_T = (y_{j,1}, ...y_{j,T}, ...\mathcal{X}_1', ..., \mathcal{X}_T')'$ be the vector containing all observations in a sample of size T.

We consider estimators of $\boldsymbol{\beta}_j^{(i)}$ that take the form

$$\hat{\boldsymbol{\beta}}_j^{(i)} = m(\mathbf{X}_T, \boldsymbol{\lambda}),$$

where $m(\mathbf{X}_T, \boldsymbol{\lambda})$ defines a machine estimator as a function of the data $\mathbf{X}_T$ and a non-negative regularization parameter vector $\boldsymbol{\lambda}$ that will be estimated using cross-validation. Denote this latter estimator $\hat{\boldsymbol{\lambda}}$ and denote the combined final estimator $\hat{\boldsymbol{\beta}}_j^{(i)}\left(\mathbf{X}_T, \hat{\boldsymbol{\lambda}}\right)$. Our main machine estimator uses Elastic Net (EN) penalties, though we have also implemented the approach with lasso and random forest. The EN estimator was the best performing, followed by lasso, while random forest performed poorly for our application. (See Online Appendix for a detailed description of the EN estimator).

The estimation of (3) is conducted in rolling subsamples, with coefficients estimated from information known at time t used predict variables $y_{j,t+h}$ in *subsequent* periods. This leads to a sequence of machine learning beliefs about $y_{j,t+h}$. Denote the coefficients estimated based on information at time t as $\hat{\boldsymbol{\beta}}_{j,t}^{(i)}$, and the time t machine learning belief about $y_{j,t+h}$ as $\mathbb{E}_t^{(i)}\left[y_{j,t+h}\right]$. This belief is defined by

$$\mathbb{E}_t^{(i)}\left[y_{j,t+h}\right] \equiv \mathcal{X}_t' \hat{\boldsymbol{\beta}}_{j,t}^{(i)}\left(\mathbf{X}_T, \hat{\boldsymbol{\lambda}}\right).$$

With this notation, forecast errors are differentially defined for the survey and machine as

$$\begin{aligned} \text{survey error}_{t+h}^{(i)} &= \mathbb{F}_t^{(i)}\left[y_{j,t+h}\right] - y_{j,t+h} \\ \text{machine error}_{t+h}^{(i)} &= \mathbb{E}_t^{(i)}\left[y_{j,t+h}\right] - y_{j,t+h}, \end{aligned}$$

with mean-squared-forecast-errors (MSEs) denoted

$$\text{survey MSE} \equiv MSE_{\mathbb{F}} = \frac{1}{P}\sum_{t=1}^{P}(\text{survey error}_{t+h})^2 \tag{4}$$

$$\text{machine MSE} \equiv MSE_{\mathbb{E}} = \frac{1}{P}\sum_{t=1}^{P}(\text{statistical error}_{t+h})^2 \tag{5}$$

where P is the length of the forecast evaluation sample.

To measure any distortions in survey expectations, we compare the forecast accuracy of the survey respondent with that of the machine. Such a comparison requires a sufficiently large number of observations on relative accuracy in order to distinguish bad luck in a random environment from a systematic mis-weighting of information. For this reason, we compare relative forecast performance, measured as the ratio of mean-squared-errors $MSE_{\mathbb{E}}/MSE_{\mathbb{F}}$, over an extended evaluation sample. If the machine benchmark consistently produces more reliable forecasts over an extended sample, we conclude that there exist systematic expectational errors, and quantify their magnitude by looking at the ratio of MSEs. Otherwise we conclude there is no systematic bias in survey expectations.

If there is evidence of systematic error, we compute a *dynamic* measure of that survey respondent's belief distortion by taking the difference between the survey forecast and the machine forecast, a time t quantity we call the "bias" for brevity. Denote the bias for forecaster i at time t as

$$bias_{j,t}^{(i)} \equiv \mathbb{F}_t^{(i)}\left[y_{j,t+h}\right] - \mathbb{E}_t^{(i)}\left[y_{j,t+h}\right]. \tag{6}$$

It is important to emphasize that $bias_{j,t}^{(i)}$ captures ex ante expectational errors, not ex post forecast errors, or "mistakes." The bias is measured as the difference between two forecasts, namely the survey forecast and a benchmark that consistently produces more reliable forecasts over an extended sample. Thus it is possible that every respondent is biased vis-a-vis the machine ex ante, even though there will always be some respondent that is "right" ex post.

Computing the statistics in (4)-(6) requires the implementation of a dynamic machine learning algorithm that processes information in real time. The full estimation and evaluation procedure involves iterating on the following steps, which are described in greater detail in the Appendix.

1. **Sample partitioning**: At time t, a prior sample of size $\widetilde{T}$ is partitioned into two subsample windows: an "in-sample" estimation subsample consisting of the first T_{IS} observations, and a hold-out "training" subsample of T_{TS} subsequent observations, i.e., $\widetilde{T} = T_{IS} + T_{TS}$.

2. **In-sample estimation**: Initial estimates of $\boldsymbol{\beta}^{(i)}$ are obtained using the EN estimator using observations $1, ..., T_{IS}$, given an arbitrary fixed (non-random) starting value for $\boldsymbol{\lambda}$. Denote this initial estimate $\boldsymbol{\beta}_{T_{IS}}^{*(i)}(\mathbf{X}_{T_{IS}}, \boldsymbol{\lambda})$, where "*" denotes the value of the estimator given an arbitrary $\boldsymbol{\lambda}$.

3. **Training and cross-validation**: The regularization parameter $\boldsymbol{\lambda}$ is estimated by minimizing mean-square loss $\mathcal{L}(\boldsymbol{\lambda}, T_{IS}, T_{TS})$ over pseudo-out-of-sample forecast errors generated from rolling regressions using only the most recent T_{IS} observations. That is, the first rolling prediction uses data from 1 to T_{IS}, the second rolling prediction uses data from 2 to $T_{IS} + 1$, etc., where

$$\mathcal{L}(\boldsymbol{\lambda}, T_{IS}, T_{TS}) \equiv \frac{1}{T_{TS} - h} \sum_{\tau=T_{IS}}^{T_{IS}+T_{TS}-h} \left(\mathcal{X}_\tau' \boldsymbol{\beta}_{j,\tau}^{*(i)}(\mathbf{X}_{T_{IS}}, \boldsymbol{\lambda}) - y_{j,\tau+h} \right)^2, \tag{7}$$

and where $\boldsymbol{\beta}_{j,\tau}^{*(i)}(\mathbf{X}_{T_{IS}}, \boldsymbol{\lambda})$ is the time τ EN estimate of $\boldsymbol{\beta}_j^{(i)}$ given $\boldsymbol{\lambda}$ and data through time τ in a sample of size T_{IS}.

4. Steps 1-3 are repeated over a grid of estimation and training sample window lengths T_{IS}^* and T_{TS}^* such that alternative partitions satisfy $T_{IS}^* + T_{TS}^* \leq \widetilde{T}$, where shorter window lengths remove consecutive observations at the start of the prior sample. The final machine estimate of $\boldsymbol{\beta}_{j,t}^{(i)}(\mathbf{X}_{\widetilde{T}}, \boldsymbol{\lambda})$ uses $\left\{\widehat{\boldsymbol{\lambda}}, \widehat{T}_{IS}, \widehat{T}_{IS}\right\} = \underset{\boldsymbol{\lambda}, T_{IS}^*, T_{TS}^*}{\operatorname{argmin}} \mathcal{L}(\boldsymbol{\lambda}, T_{IS}^*, T_{TS}^*)$ and is denoted $\widehat{\boldsymbol{\beta}}_{j,t}^{(i)}\left(\mathbf{X}_{\widetilde{T}}, \widehat{\boldsymbol{\lambda}}\right)$.

5. **Out-of-sample prediction**: The values of the regressors at time t are used to make a true out-of-sample prediction of y_{t+h}, using $\widehat{\boldsymbol{\beta}}_{j,t}^{(i)}\left(\mathbf{X}_{\widetilde{T}}, \widehat{\boldsymbol{\lambda}}\right)$, and the machine forecast error $y_{t+h} - \mathcal{X}_t' \widehat{\boldsymbol{\beta}}_{j,t}^{(i)}\left(\mathbf{X}_{\widetilde{T}}, \widehat{\boldsymbol{\lambda}}\right)$ stored.

6. **Roll forward and repeat**: The initial in-sample subperiod is rolled forward one period and uses data from $2, ..., T_{IS} + 1$ and steps 2-5 are repeated until the final out-of-sample forecast is made for $y_{j,T}$, where T is the last period of our sample.

Two points about this procedure bear noting. First, $MSE_{\mathbb{E}}$ is computed by averaging across the squared forecast errors from step 5 above over the forecast *evaluation sample,* which runs from $t = (\widetilde{T} + h)$,...,T. Belief distortions are quantified by the ratio $MSE_{\mathbb{E}}/MSE_{\mathbb{F}}$ over the evaluation sample. Second, the procedure implements not one but a series of trainings as it moves through the evaluation sample. Because each new training includes an optimized selection of estimation and training sample windows lengths, the machine can in principle adapt to a changing economic environment. This can be especially important for the estimate of the intercept, which functions as a time-varying mean estimated over optimally chosen rolling window lengths of recent past data.

Although the machine forecasting model just described is likely to perform well in "normal" times, it is not suitable for capturing extreme nonlinearities associated with times of rapid economic change, as in recessions. We therefore augment the machine algorithm so that it switches to a simpler specification when a specific recession indicator passes a threshold in real time. For this purpose we use the Treasury yield term spread. When the term spread is sufficiently low in the real time sample, the machine bases its forecasts solely on a term spread dummy indicator. The machine considers different dummy indicators that take the value 1 when the term spread at $t - 4$ is at or below some threshold, and chooses the threshold that minimizes mean-square loss in the relevant training sample immediately prior to the actual forecast.

3.3 Data

The data used for this study fall into several categories. For each category the sources and details are left to the Online Appendix. We describe each category in adumbrated form here and refer the reader to the Appendix for greater detail.

Survey Data The first data category is the survey data. We study three different surveys that ask about expectations for future inflation and aggregate economic activity: the Survey of Professional Forecasters (SPF), the University of Michigan Survey of Consumers (SOC), and the Blue Chip Survey (BC). The first covers professional forecasters in a variety of institutions, the second covers households and is designed to be representative of the U.S. population, and the third covers executives of financial firms. Data from the SPF and the SOC are publicly available; BC data were purchased and hand-coded for the earlier part of the sample.

The SPF is a quarterly survey. Respondents provide nowcasts and quarterly forecasts from one to four quarters ahead. We focus on the survey questions that ask about the level of the GDP deflator (PGDP) and about the level of real GDP. Forecasts of levels are converted to four quarter-ahead inflation and GDP growth forecasts by dividing the forecasted level by the survey respondent's nowcast. For example, forecasts of annualized inflation, denoted π_t, are computed as

$$\mathbb{F}_t^{(i)}\left[\pi_{t+h,t}\right] = (400/h) \times \ln\left(\frac{\mathbb{F}_t^{(i)}\left[P_{t+h}\right]}{\mathbb{N}_t^{(i)}\left[P_t\right]}\right).$$

The SOC asks households directly about inflation, and we use the questions on whether households expect prices to go up or down during the next twelve months to gauge their expectations about inflation. Following Curtin (2019), we take these forecasts to be most relevant for annual consumer price index (CPI) inflation, and therefore compare SOC forecasts to actual outcomes for CPI inflation. Since the SOC doesn't directly ask about GDP growth, we take the approach discussed in Curtin (2019) which is based on responses to question A7: *About a year from now, do you expect that in the country as a whole business conditions will be better, or worse than they are at present, or just about the same?* This qualitative economic forecast is converted to a point forecast for GDP growth by fitting a regression of future GDP growth data to the balance score for A7 (% respondents expect economy to improve - % expect worsen + 100) using rolling regressions and real-time GDP data.

For the BC survey, forecasters are asked to predict the average quarter over quarter percentage change in Real GDP and the GDP Price Index and the Consumer Price Index, beginning with the current quarter and extending four to five quarters into the future.

For all surveys, we align timing of survey response deadlines with real-time data, so that respondents and machine could only have used data available in real time *before* the survey deadline.

The next sections describe several large panel datasets of information that we use in our machine learning model of beliefs. We describe only the general categories of data used, and leave the lists of individual series to the Appendix.

Real Time Macro Factor Data The second data category are the real time macro data used to form real-time factors. At each forecast date, we construct a dataset of real-time quarterly macro variables observed on or before the day of the survey deadline. The real-time data are obtained from the Philadelphia Fed's Real-Time Dataset, which provides a time-series of different vintages of a macro variable for time t. The resulting real-time macro dataset, denoted $\mathcal{D}^M$, contains observations on 92 real-time macro variables. In addition to the Philadelphia Fed's real-time dataset, we include energy prices from the Bureau of Labor Statistics in $\mathcal{D}^M$. Energy prices are not revised, so they do not have multiple vintages.

The real-time macro dataset also provides observations on the left-hand-side variables about which forecasts are formed. Thus we use vintages of real-time inflation and GDP growth on the left-hand-side of (1). Following Coibion and Gorodnichenko (2015), we assume that the vintage of data respondents target when they forecast a variable is the one that is available four quarters after the period being forecast. For example, the forecast error for a survey forecast of P in 2017:Q2 that is made based on data as $t = 2016\text{:Q2}$ is computed by comparing the survey forecast $\mathbb{F}^{(i)}_{2016:\text{Q2}}\left[P_{2017:\text{Q2}}\right]$ with the actual value of $P_{2017:\text{Q2}}$ given in the 2018:Q2 vintage of the real time dataset.

Monthly Financial Factor data To take into account financial market data, we also form factors from a large panel dataset of monthly financial indicators. The dataset $\mathcal{D}^F$ uses 147 monthly financial series that include valuation ratios, growth rates of aggregate dividends and prices, default and term spreads, yields on corporate bonds of different ratings grades, yields on Treasuries and yield spreads, and a broad cross-section of industry equity returns. $\mathcal{D}^F$ also includes a group of "risk-factors" such as the three Fama and French (1993) risk factors, other risk-related portfolio returns, the momentum factor UMD_t, and the small stock value spread. We convert the monthly factors formed from the dataset $\mathcal{D}^F$ into quarterly factors by using the first month's observation for each quarter.

Daily Financial Factor Data Finally, we take into account "up-to-the-forecast" information in financial market data by using daily data on such variables up to one day before the survey respondents forecasts are due. Thus, we construct a daily financial dataset, $\mathcal{D}^D$, with series from five broad classes of financial assets: (i) commodities prices (ii) corporate risk variables (iii) equities (iv) foreign exchange, and (v) Government Securities. In total, we use 87 such series (39 commodity and futures prices, 16 corporate risk series, 9 equity series plus implied volatility, 16 government securities, and 7 foreign exchange variables).

In order to use both daily and quarterly data in our estimation, we use mixed data sampling frequency techniques. These involve taking daily indicators and converting them to quarterly factors in two steps. First, the raw daily data are used to form factors at daily frequency, G_d^D, where d denotes a business day. Second, the daily factors are converted to quarterly factors by weighting daily data in the quarter t

$$G_t^Q\left(\mathbf{w}\right) \equiv \sum_{d=1}^{N_D} w_d G^D_{N_D-d,t}.$$

Here N_D is maximum number of business days before the survey deadline in quarter t for which daily data are used. The weighting w_d function is flexibly specified as a function of a few parameters so that it can take various shapes. Typically these shapes eventually downweight more distant information but it need not do so monotonically, depending on the parameters. The parameters themselves are chosen dynamically as part of machine learning problem in order to minimize mean-squared forecast error in the hold-out training samples.

Additional Non-Factor Data A number of other non-factor variables are also include in the machine model in $\mathbf{W}'_{jt}$. These include the ith percentile's own nowcast for the variable being forecast, lags of the ith percentile's own and forecast and those of other percentiles, higher-order cross-sectional moments of the forecast observations, such as cross-sectional variance and skewness, several autoregressive lags of the left-hand-side variables, several long-term trend inflation measures, and following Hamilton (2018) measures of detrended employment and GDP.

In all, the machine model entertains a total of 68 predictor variables for inflation and 72 predictor variables for the GDP growth. The complete list of predictor variables is given in the internet Appendix. Below we refer to estimated factors with an economic name. The economic name refers to the group names given to individual series and corresponds to the individual series that generates the highest average R^2 in regressions of each series onto that estimated factor. For example, if non-farm payrolls from the Employment group has the highest average R^2 in regressions on the first common macro factor from real time macro dataset, then that factor is labeled an "Employment" factor and normalized so that it increases when non-farm payrolls increase. This gives a sense of which economic information the factor loads most heavily. The Appendix describes this procedure in greater detail.

4 Results

4.1 Preliminary Analysis

Before getting into our main findings, we begin with some preliminary analysis designed to illustrate the potential importance of certain aspects of our real time learning problem.

One such aspect is the contrast between ex ante and ex post forecasting. As an illustration of when this could be important, we consider the regressions run in Coibion and Gorodnichenko (2015) (CG) showing that mean survey forecast errors are positively predicted by ex ante mean forecast revisions. We first reproduce their findings for the SPF, which are all generated from in-sample regressions, on updated data in panel A of Table 1. As in CG, lagged forecast revisions explain next period's forecast error. Moreover, other information, e.g., lagged inflation, is unimportant in predicting mean forecast errors once the information in forecast revisions is taken into account. CG observe that these findings are consistent with the implications of theories of information frictions and stress that the implications are a property of the aggregation across individuals, not a property of the individual forecasts. The bottom panel of Table 1 runs the CG regressions out-of-sample rather than in-sample. We find that over a range of forecast evaluation subsamples using either rolling or recursive regressions, the mean-squared forecast error generated by the mean SPF survey forecast is much lower than that of a regression model that includes the information in the lagged revision of the mean forecast. Thus in contrast to the in-sample findings, in an out-of-sample context, lagged forecast revisions substantially worsen predictions of mean survey forecast errors. This result echoes a body of prior econometric evidence finding that consensus survey forecasts of inflation (mean or median) are hard to beat or even match out-of-sample with statistical models.[3]

The contradictory in-sample and out-of-sample evidence might still be compatible with information frictions. One possibility is that coefficient on the lagged forecast revision is non-zero in population, but that the sampling error is sufficiently high that what is revealed to be important with the benefit of a longer sample and hindsight is simply not apparent ex ante in a finite sample. It is, however, impossible to know whether beliefs are distorted due to information frictions or any other reason unless the econometric investigation adheres to the data availability structure survey respondents were faced with at the time they formed their forecasts. After all, even agents (such as our machine) who face no substantive information processing limitations will optimally downweight information that might appear relevant ex post if it fails to improve ex ante forecasts. We return to the question of what information, if any, is included in lagged forecast revisions in our machine learning estimation section below.

Next we turn to a second aspect of our learning problem, namely the data-rich environment. We find that predictive information found elsewhere to be important for out-of-sample forecasting in a low-dimensional setting is not necessarily important in a high-dimensional setting. As an illustration, we consider an exercise in the spirit of Chauvet and Potter (2013), who looked over a variety of low dimensional statistical models and found that a second-order autoregression often performed best. Figure 1 shows the estimated autoregressive coefficients in high versus low dimensional out-of-sample forecasts of GDP growth from one-quarter-ahead rolling regressions on predictors. The high dimensional estimation entertains a very large numbers of potential predictor variables, in the same way that our machine estimation described above does. These predictors include the two autoregressive lags. The low dimensional estimation uses the two autoregressive lags and only two additional predictors: the SPF median forecast of GDP growth four quarters ahead and the current nowcast for GDP growth (both of which are also included in the high dimensional model). The figure shows that the coefficient on the first autoregressive lag is zero in the high dimensional setting, while it is much larger than zero in the low-dimensional setting. Evidently the first autoregressive lag is unimportant once ad-

[3]For example, Ang, Bekaert, and Wei (2007), Del Negro and Eusepi (2011), Andersen, Bollerslev, Christoffersen, and Diebold (2011), Genre, Kenny, Meyler, and Timmermann (2013), and Faust and Wright (2013), conclude that professional forecasts are at the frontier of our forecasting ability. Ang et al. find that even concensus forecasts of households do better than statistical models in forecasting inflation.

ditional information is entertained. This result does not mean that the machine rarely chooses a sparse specification (indeed we report below that it often does). Instead, it indicates that it is impossible to know *which* small number of predictors are likely to be informative without the benefit of hindsight. The challenge for real time decision-making is that different pieces of information are relevant at different points in time, and forecasts eventually miss relevant information if they have not entertained large and varied datasets. Accounting for the vast array of information that could become pertinent in the future even if it wasn't in the recent past therefore requires a dynamic large-scale learning algorithm.

4.2 Forecast Comparison

We now compare the accuracy of forecasts made by the machine benchmark and the survey respondents. If the machine benchmark consistently produces more reliable forecasts over an extended sample, we take that as evidence of systematic expectational error. Otherwise we conclude there is no evidence of systematic error.

For each survey, we evaluate the relative forecast performance over the longest common sample available for all machine specifications after taking into account the different in-sample and training-sample window lengths chosen by the machine in each case. The evaluation sample spans from 1995:Q1 to 2018:Q2 for the SPF inflation and GDP growth forecasts. For SOC and BC, the evaluation samples are shorter because their survey forecasts become available at later dates. For SOC, the evaluation sample spans 1995:Q1 to 2018:Q2 for GDP growth and 1996:Q4 to 2018:Q2 for inflation. For BC, the evaluation sample spans 1997:Q1 to 2018:Q2 for GDP growth and 1997:Q3 to 2018:Q2 for inflation Tables 2 and 3 reports the out-of-sample survey $MSE_{\mathbb{F}}$ and machine $MSE_{\mathbb{E}}$ for inflation and GDP growth, respectively, for all three surveys over their respective forecast evaluation samples.

Table 2 shows that the machine model performs better than the survey forecasts of inflation for all surveys as measured by the ratio $MSE_{\mathbb{E}}/MSE_{\mathbb{F}}$, which is less than one for all percentiles, sometimes by large amounts. The machine model improves on the consensus forecasts (mean and median), which are known to be difficult to beat or even match by statistical models out-of-sample, as noted above. The magnitude of improvement is modest for the mean SPF forecast, where the ratio $MSE_{\mathbb{E}}/MSE_{\mathbb{F}}$ is 0.95. By contrast, the ratio $MSE_{\mathbb{E}}/MSE_{\mathbb{F}}$ for the median SPF forecast is 0.85. It is worth remembering that the *mean* SPF forecast is always an amalgam that does not correspond to the belief of any single respondent in the survey. It is arguably less relevant to the study of what, if any, systematic errors individuals may make when forming macroeconomic expectations. These ratios are similar for the BC survey, as shown in the last panel, where in this case $MSE_{\mathbb{E}}/MSE_{\mathbb{F}}$ is 0.84 for both the mean and the median respondent. In general, the magnitude of measured belief distortions about future inflation is much larger for SOC respondents than for the SPF and BC respondents, as shown in the middle panel. The SOC mean and median $MSE_{\mathbb{E}}/MSE_{\mathbb{F}}$ ratios are 0.58 and 0.42, respectively.

For GDP growth, Table 3 shows that machine model is also always more accurate than the survey respondents. The $MSE_{\mathbb{E}}/MSE_{\mathbb{F}}$ ratios for the mean and median SPF forecasts of GDP growth are 0.83 and 0.89, respectively, while for the BC survey they are 0.83 and 0.76, respectively. For the SOC, there is only a single forecast, denoted as if it corresponds to the "median" household. This is because the SOC forecast is constructed from the balance score for business conditions expectations, a construction that eliminates the heterogeneity (see above). The $MSE_{\mathbb{E}}/MSE_{\mathbb{F}}$ for this single SOC forecast of GDP growth is 0.74.

In summary, the results show that the machine model systematically generates better forecasts, on average, over an extended forecast evaluation sample. It does so by altering the relative weight placed on real-time information versus the survey forecast, while learning dynamically about how to do so as it moves through a forecast evaluation sample.

Given these gains in forecast accuracy, it is of interest to consider the nature of the empirical specifications chosen by the machine. We therefore close this section by reporting on the strength of the ridge and lasso penalties that are part of the dual-penalty EN estimator and chosen via cross-validation in the dynamic machine learning algorithm. Figure 2 reports a scatter plot that quantifies the strength of these penalties, with each point representing a combination of the two penalties chosen for one time period of the evaluation sample. The y-axis displays the degree of sparsity implied by the L^1 (lasso) penalty, as measured by the fraction of non-zero coefficients. The x-axis displays the degree of shrinkage implied by the L^2 (ridge) penalty, as measured by $1/\left(1+\widehat{\lambda}_{2,t}\right)$, where $\widehat{\lambda}_{2,t}$ is the estimated ridge penalty parameter for period t. The right border of the plot is the case where there is no ridge penalty at all, while the top edge of the plot is the case where there is no lasso penalty. The figure shows that machine's dynamic regularization algorithm often results in a sparse specification. In many time periods the fraction of non-zero coefficients hovers around 10% or less. But this does not happen all the time. In some periods the machine chooses very little if any sparsity, but much greater L^2shrinkage, while in other periods it chooses a lot of sparsity but little L^2 shrinkage. Occasionally, the machine chooses minimal sparsity and minimal L^2 shrinkage. This demonstrates that achieving the machine-optimal forecast calls for entertaining large amounts of information in every period, even though most of the time much of the information is associated with a coefficient that is shrunk all the way to zero. Moreover, even when sparse specifications are chosen, different sparse information sets are relevant at different points in time. Since it is impossible to know with certainty which information may be relevant ex ante, "openness" to wide-ranging and rich sources of information are vital for improving forecast accuracy over extended periods of time.

With this evidence in hand, we now explore how the estimated belief distortions vary over our sample, and which information is most mis-weighted in generating those distortions.

4.3 Dynamics of Belief Distortions

To investigate the dynamics of systematic expectational errors, we track the difference between the survey forecast and the machine forecast over the forecast evaluation samples. To this end, we report $bias_{j,t}^{(i)} \equiv \mathbb{F}_t^{(i)}[y_{j,t+h}] - \mathbb{E}_t^{(i)}[y_{j,t+h}]$ over time in several plots. Figure 3 shows biases associated with the mean and median respondents for all three surveys. The units of $bias_{j,t}^{(i)}$ are the same as the forecasts themselves and are in annual percentage points.

Figure 3 shows that systematic errors in consensus forecasts are large in some time periods, and can range between 50% and 400% of the average annual inflation or GDP growth, depending on the survey. Survey forecasts for both inflation and GDP growth oscillate between "optimism" and "pessimism." For example, for GDP growth we find extended periods of over-optimism that is especially prevalent for professional forecasters in the post-Great Recession period. For 2010:Q1-2018:Q2, the median SPF forecast of GDP growth is biased upward by 0.83% at an annual rate, or 37% of actual GDP growth during this period. The upward bias in median SPF growth expectations amounts to 20% of actual GDP growth for full evaluation sample1995:Q1-2018:Q2. These findings are quite similar for the BC survey, where the average upward bias in growth expectations amounts to 22% of actual GDP growth for 1995:Q1-2018:Q2. This suggests a robust over-optimism in SPF and BC consensus forecasts of economic growth. For the SOC, the average bias over time is close to zero even though the SOC forecast is less accurate. This happens because the SOC forecast makes systematic errors of greater magnitude that fluctuate more between optimism and pessimism.

For inflation, Figure 3 shows that mean and median expectations are biased upward (a direction we define as pessimistic) over most of the sample for the SPF and the SOC, while the BC survey exhibits an average bias that is close to zero. Nevertheless, the median forecast across all respondent types also exhibits extended periods of optimism about inflation, despite being pessimistic on average over the full sample. For example, median inflation forecasts exhibit a downward bias from 2011 to 2014 that ranges across respondent-types from –0.34% to -1.03% at an annual rate, or -19% to -47% of actual inflation during this period. Given that inflation has been declining over time, this could be interpreted as evidence of a learning process.

In summary, the estimates for consensus forecasts suggest that systematic expectational errors vary over the sample but are at times large, for example on the order of 1% or more for GDP growth in the post Great Recession period and on the order of –0.5% to -1% for annual inflation during the 2010-2014 period.

The next three figures contrast the common and heterogenous components of these belief distortions over time, breaking them out by survey. The common component is measured as the first principle component of $bias_{j,t}^{(i)}$ across all percentiles i, with heterogeneity exhibited by the distribution of $bias_{j,t}^{(i)}$ across i. Figure 4 shows the common and heterogeneous components for the SPF survey. We observe large variation in belief distortions over time that is common across SPF respondents. The optimism about economic growth in the post Great Recession subsample is present in the common component, as is a downward bias to inflation expectations for much of this same period. At the same time, there is substantial heterogeneity in beliefs, with the most optimistic and pessimistic responses differing in some time periods (typically after recessions) by 4 percent or more for GDP growth and by more than 2% for inflation. These findings are qualitatively similar for the BC survey, as shown in Figure 5. For both surveys, there are large spikes in the biases at the cusp of the 2000-2001 recession, which we discuss further below.

For the SOC, belief distortions are large and volatile over time, as shown in Figure 6. (There is only one percentile for GDP growth due to use of the balance score.) Heterogeneity in the magnitude of belief distortions for SOC respondents is enormous, especially in the period immediately after the Great Recession, where the respondent at the 95th percentile expected annual inflation four quarters ahead of 15%, while the respondent at the 5th percentile expected annual inflation of less than -5%.

The next and last figure of this section plots the forecasted values for four-quarter-ahead inflation or GDP growth along with the actual outcomes for these variables over the relevant forecast evaluation sample for each survey. Figure 7 plots the median forecast along with the actual inflation or GDP growth rate during the corresponding four quarter period. Several points about this figure bear noting.

First, from Tables 2 and 3 we observe that the forecast of the median respondent performs worse on average than the machine, but Figure 7 shows that this is clearly not true in every period. This underscores the distinction between *luck* or random error, and a systematic expectational bias.

Second, the machine typically performs better later in the sample than earlier. For all surveys, the machine has been more accurate, sometimes by large amounts, than the median respondent over the last five years of the evaluation sample, a time when expectations were biased upward for both inflation and GDP growth. For example, from 2013:Q2 to 2018:Q2, the $MSE_{\mathbb{E}}/MSE_{\mathbb{F}}$ ratio for SPF GDP growth is is 0.70, while it is 0.69 for BC. the $MSE_{\mathbb{E}}/MSE_{\mathbb{F}}$ ratio for SOC inflation over this period is 0.47, while it is 0.67 for BC inflation. What this shows is that, although respondents in the mid 1990s and early 2000s may not have had access to the same information-processing capacity that our machine model relies on today, this does not seem to confer an out-sized advantage to the machine in the early periods of our evaluation sample. This may be because there are countervailing forces that could work to the machine's disadvantage in those periods, such as the reliance by the machine on a relatively short time series of real-time quantitative data in the early recursions of the learning algorithm, and the lack of access to timely qualitative and quantitative information available to survey respondents in those periods unavailable to the machine because they weren't suitably archived.

Third, it has been observed that professional forecasters including Federal Reserve staff economists made very large forecast errors that turned out to be overly optimistic about GDP growth during the Great Recession (e.g., Gennaioli and Shleifer (2018), Chapter 2). This pattern is likewise evident in Figure 7 for all surveys studied here, where its clear that both professional forecasters and households missed by large amounts the sharp decline in economic growth during the Great Recession. At the same time, the figure shows that these large forecast errors were made by *both* the surveys and the machine model, with the machine model doing somewhat better than the SOC forecast, only slightly better than the BC forecast, and about the same but if anything slightly worse than the SPF forecast during Great Recession. Keeping in mind that the machine forecast takes into account hundreds of pieces of information including what was in financial markets in real-time, this episode underscores the point that large forecast errors *per se*, especially during sharp turning points in the economy where uncertainty is high, do not by themselves constitute evidence of systematic expectational bias.

4.4 Bias Decomposition

We now turn to an analysis of *what* information was mis-weighted in generating the distortions documented above. Recall that the time t bias is defined as the difference between the survey respondent and machine forecasts:

$$\begin{aligned} bias_{j,t+h}^{(i)} &= \mathbb{F}_{j,t+h|t}^{(i)} - \mathbb{E}_{j,t+h|t}^{(i)} \\ &= \mathbb{F}_t^{(i)}\left[y_{j,t+h}\right] - \hat{\alpha}_j - \hat{\beta}_{j\mathbb{F}}^{(i)}\mathbb{F}_t^{(i)}\left[y_{j,t+h}\right] - \hat{\mathbf{B}}_{j\mathcal{Z}}^{(i)\prime}\mathcal{Z}_{jt} \\ &= -\hat{\alpha}_j^{(i)} + \left(1-\hat{\beta}_{j\mathbb{F}}^{(i)}\right)\mathbb{F}_t^{(i)}\left[y_{j,t+h}\right] - \hat{\mathbf{B}}_{j\mathcal{Z}}^{(i)\prime}\mathcal{Z}_{jt} \end{aligned} \tag{8}$$

We are interested in the contribution of the three terms on the right-hand-side of (8). We decompose $bias_{j,t+h}^{(i)}$ into three sources of variation:

$$\begin{aligned} \text{Contribution of intercept} &: \quad -\hat{\alpha}_j^{(i)} \\ \text{Contribution of Survey Forecast:} & \quad \left(1-\hat{\beta}_{j\mathbb{F}}^{(i)}\right)\mathbb{F}_t^{(i)}\left[y_{j,t+h}\right] \\ \text{Contribution of information in } \mathcal{Z}_{j,k,t} &: \quad -\hat{B}_{j,k}^{(i)}\mathcal{Z}_{j,k,t}, \text{ for } k=1,2...K, \end{aligned} \tag{9}$$

where variables with a k subscript refer to the kth element of $\hat{\mathbf{B}}_{j\mathcal{Z}}^{(i)\prime}$ or $\mathcal{Z}_{jt}$. The sum of these three terms equals 100% of $bias_{j,t+h}^{(i)}$. This decomposition gives an indication of *which* information is most mis-weighted by the survey respondent, and by how much. The intercept term $\hat{\alpha}_j^{(i)}$ changes over the evaluation sample through the dynamic algorithm and is akin to a time-varying conditional mean applied to the most recent rolling subsample window. In the discussion below we denote this "rolling mean" with a t subscript, i.e., $\hat{\alpha}_{j,t.}^{(i)}$ For the same reason, the estimates $\hat{\beta}_{j\mathbb{F}}^{(i)}$ and $\hat{\mathbf{B}}_{j\mathcal{Z}}^{(i)\prime}$ also vary over the evaluation sample and are therefore sometimes denoted with a t subscript.

To interpret the machine weighting of information vis-a-vis the forecast, it is useful to consider the magnitude and signs of the coefficients in the components above. For example, if $\hat{\beta}_{j\mathbb{F}}^{(i)} < 1$, this implies that the machine improves forecasts by downweighting the survey forecast in favor of giving more weight to other information. Thus an estimate of $\hat{\beta}_{j\mathbb{F}}^{(i)} < 1$ implies that the respondent at the ith percentile *over*-weighted on her own forecast, and in that sense is overconfident. On the other hand, if $\hat{\beta}_{j\mathbb{F}}^{(i)} > 1$, this implies that the machine improves forecasts by giving greater weight the survey forecast than the implicit weight given by the respondent to her own forecast relative to other information, suggesting respondent underconfidence.

For the information variables and the rolling mean, any estimate of $\widehat{B}_{j,k}^{(i)} \neq 0$ or $\hat{\alpha}_{j,t}^{(i)} \neq 0$ and indicates that the machine improved forecasts by giving greater absolute weight to $\mathcal{Z}_{j,k,t}$ or $\hat{\alpha}_{j,t}^{(i)}$ compared to the respondent's implicit weight of zero conditional on her own forecast. Thus we say that estimates with $\widehat{B}_{j,k}^{(i)} \neq 0$ or $\hat{\alpha}_{j,t}^{(i)} \neq 0$ imply that the respondent *under*-weighted these sources of information. We summarize the respondent over- versus under-weighting as follows:

$$\begin{aligned} \text{Over-weight} \quad &: \quad \hat{\beta}_{j\mathbb{F}}^{(i)} < 1; \\ \text{Under-weight} \quad &: \quad \hat{\beta}_{j\mathbb{F}}^{(i)} > 1;\ \widehat{B}_{j,k}^{(i)} \neq 0;\ \hat{\alpha}_{j,t}^{(i)} \neq 0. \end{aligned}$$

The next set of figures reports the contribution over time to the median bias, $bias_{j,t+h}^{(50)}$, of the three components in (9), i.e., the figure reports the (negative) of the intercept, $-\hat{\alpha}_{j,t}^{(i)}$, the survey forecast $\left(1 - \hat{\beta}_{j\mathbb{F},t}^{(i)}\right) \mathbb{F}_t^{(i)} \left[y_{j,t+h}\right]$, and the most important information contributors $\mathcal{Z}_{j,k,t}$ to the information variable component, $-\hat{B}_{j,k,t}^{(i)} \mathcal{Z}_{j,k,t}$. Since there are many $\mathcal{Z}_{j,k,t}$ that may be important at different t, we report the contributions of only those information variables that have the largest average absolute impact on the bias, as measured by the absolute sum of the information variable's contributions over the evaluation sample, i.e., $\sum_t \left| -\widehat{B}_{j,k,t}^{(50)} \mathcal{Z}_{j,k,t} \right|$, where $\widehat{B}_{j,k,t}^{(50)}$ denotes the rolling, real-time estimate of $B_{j,k}^{(50)}$ based on data available through period t of the evaluation sample. The solid lines in Figures 8- 13 report the total bias, $bias_{j,t+h}^{(50)}$. The contributions themselves are reported as bar charts, where a bar is above zero if the relevant contribution of the component in (9) is positive, and below zero if it is negative. For example, for the contribution of the survey forecast, a bar is above zero if $\left(1 - \hat{\beta}_{j\mathbb{F}}^{(50)}\right) \mathbb{F}_t^{(50)} \left[y_{j,t+h}\right]$ is positive at time t, which means that the product of $\left(1 - \hat{\beta}_{j\mathbb{F}}^{(i)}\right)$ and $\mathbb{F}_t^{(i)} \left[y_{j,t+h}\right]$ had an *up*ward effect on the overall bias. Conversely, a bar is below zero if $\left(1 - \hat{\beta}_{j\mathbb{F}}^{(50)}\right) \mathbb{F}_t^{(50)} \left[y_{j,t+h}\right]$ is negative, which means that the product had a *down*ward effect on the overall bias. Finally, the color of the bars indicates whether the median survey respondent gave too much or too little weight to her own forecast. A red bar indicates that she over-weighted her own forecast, while a blue bar indicates that she under-weighted. For the intercept and information variables, any bar with a non-zero height indicates that the respondent under-weighted that information. Figures 8-10 exhibit this information for survey expectations of inflation, while Figures 11-13 do so for survey expectations of GDP growth.

A key finding evident from all of these figures is that survey respondents almost always place too much weight on their own forecast relative to other information, and are in that sense overconfident. This happens for all surveys, for both inflation and GDP growth, and for most time periods. This is exhibited in the Figures by the frequent red-colored bars in the Survey Forecast panels. Much of the time, but not always, this overconfidence tends to happen when the survey forecast contributes positively to the measured bias. The length of the bars indicates that the respondent's over-weighting of her own forecast contributes in most cases to quantitatively large distortions in macro expectations.

For example, Figure 8 indicates that the median SPF respondent's forecast of four-quarter-ahead inflation contributed 4%–or more than 100%–to the total upward bias in inflation expectations during several periods right after the Great Recession. (Recessions are shown in the figure by light grey shaded bars.) This can be observed in the Survey Forecast subplot by the tall above-zero bars. That these bars are all red indicates that the machine improved forecasts by greatly downweighted the survey forecast in these periods in favor of placing more absolute weight on the rolling mean, on a measure of long-run inflation, on the two-period lagged value of the median SPF inflation forecast, and on daily financial factors related to Treasuries and corporate risk. The findings are qualitatively similar for the BC and SOC surveys of inflation, though the information variables most mis-weighted are different across these surveys. In general finding across all surveys, however, is that the survey respondent's own forecast is the single most important contributor quantitatively to $bias^{(50)}_{j,t+h}$, as indicated by the height of the bars in the survey forecast subpanel for each case. Moreover in each case, the bars are red, indicating that survey respondents placed too much weight on their own forecast compared to the machine-efficient benchmark. Rarely, if ever, do the respondent's under-weight their own forecast.

Figures 11-13 show that these conclusions are even more true for the median survey expectations of economic growth. For all three surveys, the median respondent's over-weighting of her own forecast is the most important quantitative contributor to the excessive optimism about GDP growth during the last several years of our sample. The tall red bars indicate that the machine greatly downweighted the survey forecasts in these periods in favor of placing more absolute weight, primarily in this case, on the rolling mean, though other respondent's forecasts of growth last period are also given non-zero weight by the machine.

We close this section by asking whether revisions in survey forecast are an important contributor to expectational biases. We do this by running the following machine version of the CG regressions, which use the mean rather than median SPF inflation forecast $\mathbb{F}^{(\mu)}$, with forecast errors on the left-hand-side:

$$\underbrace{\pi_{j,t+3} - \mathbb{F}_t^{(\mu)}\left[\pi_{j,t+3}\right]}_{\text{forecast error}} = \alpha_\pi^{(\mu)} + \beta_{\pi\mathrm{FR}}^{(\mu)}\left(\underbrace{\mathbb{F}_t^{(\mu)}\left[\pi_{t+3}\right] - \mathbb{F}_{t-1}^{(\mu)}\left[\pi_{t+3}\right]}_{\text{forecast revision}}\right) + \mathbf{B}_{\pi\mathcal{Z}}^{(\mu)\prime}\mathcal{Z}_{\pi t} + \epsilon_{\pi t+h}. \tag{10}$$

The machine estimates differs from the CG estimation for three reasons. First, the machine forecasts are made out-of-sample. Second, the machine entertains the large-scale information-set $\mathcal{Z}_{\pi t}$ as additional potential predictor variables, while CG use only the forecast revision. Third, the machine uses the EN estimator while CG use least squares. Denote the CG estimate of the coefficient on forecast revisions from this univariate, in-sample least squares regression as $\beta_{\pi\mathrm{CG}}^{(\mu)}$.

Figure 14 reports the coefficients $\beta_{j\mathrm{FR}}^{(\mu)}$ obtained from estimating (10) using the machine learning algorithm described above. Since the estimation is repeated on rolling samples using real time information up to time t, the figure reports the entire time-series of estimates $\widehat{\beta}_{j\mathrm{FR},t}^{(\mu)}$ using a bar chart, where the height of the bar indicates the magnitude of $\widehat{\beta}_{j\mathrm{FR},t}^{(\mu)}$ and the time period t of the forecast evaluation sample 1995:Q1-2018:Q2 is given on the x-axis. Time periods τ for which there is no bar displayed indicate $\widehat{\beta}_{j\mathrm{FR},\tau}^{(\mu)} = 0$. For comparison, in-sample estimates $\widehat{\beta}_{\pi\mathrm{CG}}^{(\mu)}$ from the CG least squares regressions are shown as separate horizontal lines, one for each of three estimation samples: 1969:Q1-2014:Q4 (CG sample), 1969:Q1-2018:Q2 (our full sample) and 1995:Q1-2018:Q2 (our machine forecast evaluation sample). The horizontal lines for $\widehat{\beta}_{\pi\mathrm{CG}}^{(\mu)}$ over the first two of these samples lie almost on top of one another, and are close to 1.2, while that for the shorter more recent sample are smaller by half. By contrast, the machine estimates $\widehat{\beta}_{j\mathrm{FR},t}^{(\mu)}$ are always much smaller than the univariate, in-sample least squares estimates $\widehat{\beta}_{\pi\mathrm{CG}}^{(\mu)}$ when those are obtained using the two longer subsamples, and they only match or exceed the half-as-large value obtained using the 1995:Q1-2018:Q2 sample in one time period. Instead, the coefficients on forecast revisions are typically shrunk to zero by the machine algorithm in favor of placing greater absolute weight on other pieces of information contained in $\mathcal{Z}_{\pi t}$ or $\widehat{\alpha}_{\pi,t}^{(\mu)}$. The coefficient $\widehat{\beta}_{j\mathrm{FR},t}^{(\mu)}$ is non-zero in only 6 out of 94 quarters over the evaluation sample. These findings do not indicate an important role for ex ante revisions in the average forecast in predicting average ex post forecast errors, as would be indicative of models with information rigidities.

4.5 Do Belief Distortions Matter for Macroeconomic Fluctuations?

What is the dynamic relationship between our measured expectational errors and macroeconomic activity? We use vector autoregressions (VARs) to investigate this question. To do so we first construct indexes of the common factor component in our measured biases and then consider two vector autoregressions (VARs) that separately use a different index of our measured belief distortions. One index, which we denote $\overline{bias}_t^{\pi}$, is constructed as the first common factor, measured as the first principle component (PC), of inflation biases $bias_{\pi,t+h}^{(i)}$ across all surveys and all percentiles i of each survey. The other index is constructed analogously to measure belief distortions in GDP growth and is denoted $\overline{bias}_t^{\Delta y}$. Since the PCs and their factor loadings Λ are not separately identifiable, we use a standard normalization to pin down the magnitudes of Λ and normalize the signs of $\overline{bias}_t^{\pi}$ and $\overline{bias}_t^{\Delta y}$so that they are positively correlated with the average median bias across surveys for π and Δy, respectively.[4] Thus an increase in $\overline{bias}_t^{\pi}$ corresponds to an increase in pessimism about inflation by the average median respondent, while an increase in $\overline{bias}_t^{\Delta y}$corresponds to an increase in optimism about economic growth by the average median respondent.

A question arises as to which variables to include in the VARs. Given the relatively short evaluation samples, we cannot entertain too many variables or too many lags.[5] We use a one-lag VARs but choose a range of variable types. To study impulse responses and variance decompositions with respect to a shock in these bias index, the covariance matrix of VAR residuals is orthogonalized using a Cholesky decomposition with variables ordered as listed below, with bias index placed last. This placement is conservative for assessing the relation of the expectational errors to macro fluctuations, since it attributes all the contemporaneous comovement between the bias index and macroeconomic indicators to shocks in the other variables. The variables in the VAR are

$$\begin{bmatrix} \log\text{(Real GDP)} \\ \log\text{(GDP Deflator)} \\ \log\text{(Real Investment)} \\ \log\text{(Real Wage)} \\ \log\text{(S\&P500)} \\ \text{federal funds rate} \\ \overline{bias}_{t+h}^{x} \end{bmatrix} \tag{VAR}$$

where $\overline{bias}_{t+h}^{x}$ is either $\overline{bias}_{t+h}^{\Delta y}$ or $\overline{bias}_{t+h}^{\pi}$. With respect to the impulse responses and variance decompositions reported below, a "shock" to bias is a movement in belief distortions that is contemporaneously uncorrelated to the aggregate economic state, as measured by the above

[4]The loadings are normalized by $(\Lambda'\Lambda)/N = \mathbf{I}_q$ where N is the number of $bias^{(i)}$ series over which common factors are formred and q is the number of common factors.

[5]For the SPF, the evaluation sample spans the periods1995:Q1 to 2018:Q2. As explained above, for the BC and SOC, observations on the bias starts a few periods after 1995:Q1. Thus we add those observations in when they become available and use only the SPF in the first few years of the 1995:Q1 to 2018:Q2 sample.

non-bias variables. The VAR is estimated with standard Bayesian methods under flat priors.

How quantitatively important are time-varying belief distortions and to what extent are they correlated with macroeconomic fluctuations? For results using the inflation bias index, it is instructive to compare three different cases. Figure 15 reports the dynamic responses using $\overline{bias}_t^{\pi}$ when this index is constructed as the first principle component of inflation biases across all percentiles in the SPF and BC surveys. This figure shows that a positive innovation to $\overline{bias}_t^{\pi}$ (indicating more pessimism about inflation) operates like a cost-push shock, driving up the real wage, but driving down prices, real investment, and real GDP. The effects on the real wage are large and persist for over five years. On the other hand, perhaps because this shock drives down prices, the effects on real GDP are smaller and more transitory. The results are quite different, however, when $\overline{bias}_t^{\pi}$ is constructed as the first principle component of the inflation biases across all percentiles of the SOC by itself, as seen in Figure 16. In this case the error bands are much wider and the results inconclusive, suggesting that biases in household-level inflation expectations exhibit little reliable relation to aggregate economic activity, in contrast to the biases in professional and corporate executive expectations. This suggests that household-level expectational errors, which are far more heterogeneous and are based on far less accurate forecasts, are more "noise" than "news." Figure 17 reports the dynamic responses using $\overline{bias}_t^{\pi}$ when this index is constructed as the first principle component of inflation biases across all percentiles of SPF and BC and the *median* of SOC. In this case, the responses are virtually identical to those in Figure 15, which use only the SPF and BC surveys.

Figure 18 shows that a positive innovation to $\overline{bias}_t^{\Delta y}$ (indicating more optimism about economic growth) leads to a sizable and protracted increase in real activity, in the price level, in the real wage, and in the stock market.[6] It is important to note that these results are specific to innovations in the systematic expectational *errors* survey respondents make about future GDP growth, and not to their expectations *per se*. Indeed, we find that a positive innovation in an index of GDP growth *expectations* has very different effects from those of the bias index $\overline{bias}_t^{\Delta y}$ and is not associated with a boom in economic activity. This may be observed from Figure 19, which shows the dynamic responses to innovations in an index of survey expectations of GDP growth, constructed as the first PC across all surveys and all percentiles of GDP growth expectations and denoted $\overline{\mathbb{F}}_t^{\Delta y}$. For the VAR used to generate this figure, we replace $\overline{bias}_t^{\Delta y}$ with $\overline{\mathbb{F}}_t^{\Delta y}$ in (VAR). In contrast to the responses to innovations in $\overline{bias}_t^{\Delta y}$, positive innovations in $\overline{\mathbb{F}}_t^{\Delta y}$ (indicating higher expected economic growth by the average median respondent) are associated with a decrease rather than an increase in real GDP, in the stock market, and real investment, though the credible sets for real investment response are wide and include both

[6] This index is constructed as the first principle component of the bias in SPF, BC, and SOC. Recall that we only have one forecast for the SOC, since this is constructed from the balance index score, hence the enormous heterogeneity in the forecast accuracy and bias that is present for household-level inflation forecasts is not present for the GDP forecasts.

positive and negative responses.

To study the quantitative importance of the bias shocks for macroeconomic fluctuations, Table 4 reports variance decompositions of the VAR variables, over several VAR forecast horizons. Specifically, we report the fraction of forecast error variance that is explained by shocks to $\overline{bias}_{t+h}^{\Delta y}$ or $\overline{bias}_{t+h}^{\pi}$ with the variables again ordered as above. We use k in this table denote the VAR forecast horizon and use "max k" to denote the forecast horizon k for which a shock explains the maximum fraction of forecast error variance. The table also reports the fraction of forecast error variance explained by shocks to the federal funds rate, which we discuss below.

Table 4 shows that innovations to the GDP growth bias index account for up to 10%, 8% and 3.2% of the forecast error variance in GDP growth, inflation, and the stock market, respectively, depending on the VAR forecast horizon. Although these magnitudes are relatively modest in absolute terms, it is worth forming a basis for comparison. Over the same sample, innovations to the federal funds rate (a common proxy for unanticipated shifts in monetary policy) explain (at most) 7%, 5.5%, and 1.4% of the forecast error variance in these same variables, despite the federal funds rate being placed ahead of the bias index in the VAR. On the other hand, the GDP growth bias index accounts for at most 7.3% and 5.7% of the forecast error variance in the real wage and real investment, respectively, compared to 7.6% and 11.7% for the federal funds rate. Overall, innovations in the inflation bias index account for less of the forecast error variance of the other VAR variables, but contributions are in the same ballpark as innovations in the federal funds rate. That the effects for both indexes are comparable to or in some cases quantitatively more important than those for the federal funds rate is consistent with the view that expectational errors have non-trivial implications for aggregate economic activity.

5 Conclusion

This paper provides new measures of systematic expectational errors in survey responses and relates them to macroeconomic activity. Biases in inflation expectations for the median respondent of all types are on average too high over our evaluation sample, a direction we refer to as "pessimistic." By contrast, biases in expectations of economic growth are "optimistic" on average–i.e., too high–for the median respondent among professional forecasters and corporate executives, while they are very slightly pessimistic for households. But these averages mask large variation over time in the median respondent's bias, as well across respondents at any given point in time. A pervasive finding across all surveys is that respondents place too much weight on their own forecast relative to other information, and are in that sense overconfident.

We find that fluctuations in belief distortions exhibit important dynamic relations with the macroeconomy. A positive innovation to an index of inflation bias (indicating an increase in pessimism) operates much like a cost-push shock, driving up the real wage and driving down real investment, real GDP, and the price level. By contrast, a positive innovation to a GDP growth bias index (indicating an increase in optimism) has the opposite effect and leads to a sizable and more protracted increase in real activity, the price level, and also the stock market, while the real wage declines. Innovations in GDP growth expectations, as opposed to the biases in those expectations, lead to very different effects and are not associated with an increase in real activity.

References

Adam, K., A. Marcet, and J. Beutel (2017): "Stock price booms and expected capital gains," *American Economic Review*, 107(8), 2352–2408.

Afrouzi, H., L. Veldkamp, et al. (2019): "Biased Inflation Forecasts," in *2019 Meeting Papers*, no. 894. Society for Economic Dynamics.

Amromin, G., and S. A. Sharpe (2014): "From the horse's mouth: Economic conditions and investor expectations of risk and return," *Management Science*, 60(4), 845–866.

Andersen, T. G., T. Bollerslev, P. F. Christoffersen, and F. X. Diebold (2011): "Forecast Combinations," in *Oxford Handbook of Economic Forecasting*, ed. by M. P. Clements, and D. F. Hendry, pp. 355–389. Oxford University Press, 198 Madison Avenue, New York, NY 10016.

Andreou, E., E. Ghysels, and A. Kourtellos (2013): "Should macroeconomic forecasters use daily financial data and how?," *Journal of Business & Economic Statistics*, 31(2), 240–251.

Ang, A., G. Bekaert, and M. Wei (2007): "Do macro variables, asset markets, or surveys forecast inflation better?," *Journal of monetary Economics*, 54(4), 1163–1212.

Angeletos, G.-M., F. Collard, and H. Dellas (2018): "Quantifying Confidence," *Econometrica*, 86(5), 1689–1726.

Angeletos, G.-M., and J. La'O (2013): "Sentiments," *Econometrica*, 81(2), 739–779.

Bacchetta, P., E. Mertens, and E. Van Wincoop (2009): "Predictability in financial markets: What do survey expectations tell us?," *Journal of International Money and Finance*, 28(3), 406–426.

Bai, J., and S. Ng (2002): "Determining the Number of Factors in Approximate Factor Models," *Econometrica*, 70(1), 191–221.

——— (2006): "Confidence Intervals for Diffusion Index Forecasts and Inference for Factor-Augmented Regressions," *Econometrica*, 74(4), 1133–50.

Barberis, N., R. Greenwood, L. Jin, and A. Shleifer (2015): "X-CAPM: An extrapolative capital asset pricing model," *Journal of financial economics*, 115(1), 1–24.

Barberis, N., A. Shleifer, and R. W. Vishny (1998): "A Model of Investor Sentiment," *Journal of Financial Economics*, 49(3), 307–43.

Ben-David, I., J. R. Graham, and C. R. Harvey (2013): "Managerial miscalibration," *The Quarterly Journal of Economics*, 128(4), 1547–1584.

Bhandari, A., J. Borovicka, and P. Ho (2019): "Survey data and subjective beliefs in business cycle models," FRB Richmond Working Paper.

Bianchi, F., C. Ilut, and M. Schneider (2017): "Uncertainty Shocks, Asset Supply and Pricing Over the Business Cycle," *The Review of Economic Studies,* forthcoming.

Bordalo, P., N. Gennaioli, R. L. Porta, and A. Shleifer (2019): "Diagnostic expectations and stock returns," *The Journal of Finance*, 74(6), 2839–2874.

Bordalo, P., N. Gennaioli, and A. Shleifer (2018): "Diagnostic expectations and credit cycles," *The Journal of Finance*, 73(1), 199–227.

Bouchaud, J.-P., P. Krueger, A. Landier, and D. Thesmar (2019): "Sticky expectations and the profitability anomaly," *The Journal of Finance*, 74(2), 639–674.

Chauvet, M., and S. Potter (2013): "Forecasting output," in *Handbook of Economic Forecasting*, vol. 2, pp. 141–194. Elsevier.

Coibion, O., and Y. Gorodnichenko (2012): "What can survey forecasts tell us about information rigidities?," *Journal of Political Economy*, 120(1), 116–159.

——— (2015): "Information rigidity and the expectations formation process: A simple framework and new facts," *American Economic Review*, 105(8), 2644–78.

Curtin, R. T. (2019): *Consumer Expectations: Micro Foundations and Macro Impact*. Cambridge University Press.

De Bondt, W. F., and R. H. Thaler (1990): "Do security analysts overreact?," *The American Economic Review*, pp. 52–57.

De Long, J. B., A. Shleifer, L. H. Summers, and R. J. Waldmann (1990): "Positive feedback investment strategies and destabilizing rational speculation," *the Journal of Finance*, 45(2), 379–395.

Del Negro, M., and S. Eusepi (2011): "Fitting observed inflation expectations," *Journal of Economic Dynamics and control*, 35(12), 2105–2131.

Epstein, L. G., and M. Schneider (2010): "Ambiguity and asset markets," *Annual Review of Financial Economics*, 2, 315–346.

Fama, E. F., and K. R. French (1992): "The Cross-Section of Expected Returns," *Journal of Finance*, 47(2), 427–65.

——— (1993): "Common Risk Factors in the Returns on Stocks and Bonds," *Journal of Financial Economics*, 33, 3–56.

Faust, J., and J. H. Wright (2013): "Forecasting inflation," in *Handbook of economic forecasting*, vol. 2, pp. 2–56. Elsevier.

Fuhrer, J. C. (2018): "Intrinsic expectations persistence: Evidence from professional and household survey expectations," FRB of Boston Working Paper.

Gennaioli, N., Y. Ma, and A. Shleifer (2016): "Expectations and investment," *NBER Macroeconomics Annual*, 30(1), 379–431.

Gennaioli, N., and A. Shleifer (2018): *A crisis of beliefs: Investor psychology and financial fragility*. Princeton University Press.

Genre, V., G. Kenny, A. Meyler, and A. Timmermann (2013): "Combining expert forecasts: Can anything beat the simple average?," *International Journal of Forecasting*, 29(1), 108–121.

Ghysels, E., A. Sinko, and R. Valkanov (2007): "MIDAS regressions: Further results and new directions," *Econometric Reviews*, 26(1), 53–90.

Greenwood, R., and S. G. Hanson (2015): "Waves in ship prices and investment," *The Quarterly Journal of Economics*, 130(1), 55–109.

Greenwood, R., and A. Shleifer (2014): "Expectations of returns and expected returns," *The Review of Financial Studies*, 27(3), 714–746.

Hamilton, J. D. (2018): "Why you should never use the Hodrick-Prescott filter," *Review of Economics and Statistics*, 100(5), 831–843.

Hansen, L. P., and T. J. Sargent (2008): *Robustness*. Princeton university press.

Huang, D., F. Jiang, and G. Tong (2017): "Real time macro factors in bond risk premium," Unpublished paper, Lee kong Chian School of Business.

Ilut, C. L., and H. Saijo (2020): "Learning, confidence, and business cycles," *Journal of Monetary Economics*.

Ilut, C. L., and M. Schneider (2015): "Ambiguous Business Cycles," *American Economic Review,* forthcoming.

Jurado, K., S. C. Ludvigson, and S. Ng (2015): "Measuring Uncertainty," *The American Economic Review*, 105(3), 117–1216.

Khaw, M. W., L. Stevens, and M. Woodford (2017): "Discrete adjustment to a changing environment: Experimental evidence," *Journal of Monetary Economics*, 91, 88–103.

Ludvigson, S. C., S. Ma, and S. Ng (2019): "Uncertainty and Business Cycles: Exogenous Impulse or Endogenous Response?," *American Economic Journal: Macroeconomics,* forthcoming, http://www.econ.nyu.edu/user/ludvigsons/ucc.pdf.

Ludvigson, S. C., and S. Ng (2007): "The Empirical Risk-Return Relation: A Factor Analysis Approach," *Journal of Financial Economics*, 83(1), 171–222.

——— (2009): "Macro Factors in Bond Risk Premia," *The Review of Financial Studies*, 22(12), 5027–67.

Malmendier, U., and S. Nagel (2011): "Depression babies: do macroeconomic experiences affect risk taking?," *The Quarterly Journal of Economics*, 126(1), 373–416.

——— (2015): "Learning from inflation experiences," *The Quarterly Journal of Economics*, 131(1), 53–87.

Martin, I., and S. Nagel (2019): "Market Efficiency in the Age of Big Data," Unpublished manuscript, University of Chicago Booth.

Milani, F. (2011): "Expectation shocks and learning as drivers of the business cycle," *The Economic Journal*, 121(552), 379–401.

——— (2017): "Sentiment and the U.S. business cycle," *Journal of Economic Dynamics and Control*, 82, 289–311.

Woodford, M. (2013): "Macroeconomic analysis without the rational expectations hypothesis," *Annu. Rev. Econ.*, 5(1), 303–346.

Table 1: **CG Regressions of Forecast Errors on Forecast Revisions**

Panel A: **In-sample Regressions (CG Sample)**		
Regression: $\pi_{t+3} - \mathbb{F}_t^{(\mu)}[\pi_{t+3}] = \alpha^{(\mu)} + \beta^{(\mu)}\left(\mathbb{F}_t^{(\mu)}[\pi_{t+3}] - \mathbb{F}_{t-1}^{(\mu)}[\pi_{t+3}]\right) + \delta\pi_{t+2,t-1} + \epsilon_t$		
Constant	0.001	-0.077
t-stat	(0.005)	(-0.442)
$\mathbb{F}_t[\pi_{t+3,t}] - \mathbb{F}_{t-1}[\pi_{t+3,t}]$	1.194**	1.141**
t-stat	(2.496)	(2.560)
$\pi_{t+2,t-1}$		0.021
t-stat		(0.435)
$\bar{R}^2$	0.195	0.197

Panel B: **Out-of-sample Regressions**		
Regression: $\pi_{t+3} - \mathbb{F}_t^{(\mu)}[\pi_{t+3}] = \alpha^{(\mu)} + \beta^{(\mu)}\left(\mathbb{F}_t^{(\mu)}[\pi_{t+3}] - \mathbb{F}_{t-1}^{(\mu)}[\pi_{t+3}]\right) + \epsilon_{t+3}$		
Method	Forecast Sample	$\text{MSE}_{CG}/\text{MSE}_{\mathbb{F}}$
Rolling 5 years	1975:Q4 - 2018:Q2	1.38
Rolling 10 years	1980:Q4 - 2018:Q2	1.29
Rolling 20 years	1990:Q4 - 2018:Q2	1.31
Recursive 5 years	1975:Q4 - 2018:Q2	1.69
Recursive 10 years	1980:Q4 - 2018:Q2	1.60
Recursive 20 years	1990:Q4 - 2018:Q2	1.33

In-sample versus out-of-sample regressions using CG specification. Panel A reports the in-sample results over the sample used in Coibion and Gorodnichenko (2015) (CG), 1969:Q1 to 2014:Q4. Newey-West corrected t-statistics with lags = 4 are reported in parenthesis. Panel B reports the ratio of out-of sample mean-squared-error (MSE) of the CG model forecast to that for the survey forecast computed using different rolling or recursive estimation windows. The MSE for the CG model averages the (square of the) forecast errors $\pi_{t+3} - \widehat{\pi}_{t+3}^{(\mu)}$, where $\widehat{\pi}_{t+3}^{(\mu)} = \widehat{\alpha}_t^{(\mu)} + \left(1 + \widehat{\beta}_t^{(\mu)}\right)\mathbb{F}_t^{(\mu)}[\pi_{t+3}] - \widehat{\beta}_t^{(\mu)}\mathbb{F}_{t-1}^{(\mu)}[\pi_{t+3}]$. In both panels, the regression estimation uses the latest vintage of inflation in real time and, following CG, computes forecast errors with real-time data available four quarters after the period being forecast. Annual inflation is defined as $\pi_{t+3,t} = \frac{P_t}{P_{t-1}} \times \frac{P_{t+1}}{P_t} \times \frac{P_{t+2}}{P_{t+1}} \times \frac{P_{t+3}}{P_{t+2}}$, and $\mathbb{F}_t[\pi_{t+3,t}]$ is the mean forecast of annual inflation as of time t from the Survey of Professional Forecasters (SPF). The sample of Panel B spans the period 1969:Q1 - 2018:Q2. *sig. at 10%. **sig. at 5%. ***sig. at 1%.

Table 2: **Machine Learning versus Survey Forecasts of Inflation**

ML: $y_{j,t+h} = \alpha_j^{(i)} + \beta_{j\mathbb{F}}^{(i)} \mathbb{F}_t^{(i)} [y_{j,t+h}] + \mathbf{B}_{j\mathcal{Z}}^{(i)} \mathcal{Z}_{jt} + \epsilon_{jt+h}$

Survey of Professional Forecasters (SPF)					
Percentile	Median	Mean	5th	10th	20th
$MSE_{\mathbb{E}}$	0.38	0.42	0.51	0.43	0.40
$MSE_{\mathbb{F}}$	0.45	0.44	0.90	0.58	0.48
$MSE_{\mathbb{E}}/MSE_{\mathbb{F}}$	0.85	0.95	0.56	0.74	0.83
	25th	30th	40th	60th	70th
$MSE_{\mathbb{E}}$	0.41	0.40	0.39	0.36	0.39
$MSE_{\mathbb{F}}$	0.45	0.45	0.44	0.49	0.56
$MSE_{\mathbb{E}}/MSE_{\mathbb{F}}$	0.90	0.88	0.89	0.74	0.70
	75th	80th	90th	95th	
$MSE_{\mathbb{E}}$	0.41	0.41	0.53	0.65	
$MSE_{\mathbb{F}}$	0.61	0.70	0.96	1.36	
$MSE_{\mathbb{E}}/MSE_{\mathbb{F}}$	0.67	0.59	0.55	0.47	
Michigan Survey of Consumers (SOC)					
Percentile	Median	Mean	5th	10th	20th
$MSE_{\mathbb{E}}$	1.64	1.97	3.25	2.28	1.76
$MSE_{\mathbb{F}}$	2.84	4.65	15.11	8.27	3.87
$MSE_{\mathbb{E}}/MSE_{\mathbb{F}}$	0.58	0.42	0.22	0.28	0.46
	25th	30th	40th	60th	70th
$MSE_{\mathbb{E}}$	1.83	1.76	1.50	1.75	1.70
$MSE_{\mathbb{F}}$	3.16	2.62	2.30	4.69	8.26
$MSE_{\mathbb{E}}/MSE_{\mathbb{F}}$	0.58	0.67	0.65	0.37	0.21
	75th	80th	90th	95th	
$MSE_{\mathbb{E}}$	1.65	1.83	2.51	2.72	
$MSE_{\mathbb{F}}$	10.62	15.05	47.03	84.92	
$MSE_{\mathbb{E}}/MSE_{\mathbb{F}}$	0.16	0.12	0.05	0.03	
Blue Chip Financial Forecasts (BC)					
Percentile	Median	Mean	5th	10th	20th
$MSE_{\mathbb{E}}$	0.41	0.40	0.48	0.40	0.43
$MSE_{\mathbb{F}}$	0.49	0.48	0.83	0.67	0.51
$MSE_{\mathbb{E}}/MSE_{\mathbb{F}}$	0.84	0.84	0.58	0.60	0.85
	25th	30th	40th	60th	70th
$MSE_{\mathbb{E}}$	0.42	0.41	0.43	0.40	0.39
$MSE_{\mathbb{F}}$	0.49	0.47	0.47	0.51	0.56
$MSE_{\mathbb{E}}/MSE_{\mathbb{F}}$	0.85	0.86	0.91	0.78	0.69
	75th	80th	90th	95th	
$MSE_{\mathbb{E}}$	0.39	0.38	0.42	0.44	
$MSE_{\mathbb{F}}$	0.59	0.65	0.87	1.14	
$MSE_{\mathbb{E}}/MSE_{\mathbb{F}}$	0.65	0.59	0.48	0.38	

Machine v.s. survey mean-square-forecast errors for inflation. $MSE_{\mathbb{E}}$ and $MSE_{\mathbb{F}}$ denote the machine learning and survey mean-squared-forecast-errors, respectively, computed for 4-quarter-ahead forecasts and averaged over the evaluation period. The evaluation period for the Survey of Professional Forecasters (SPF) is 1995:Q1 to 2018:Q2; for the Michigan Survey of Consumers (SOC) is 1996:Q4 to 2018:Q2; and for the Bluechip (BC)survey is 1997:Q3 to 2018:Q2. The full estimation sample spans the periods 1969:Q3 to 2018:Q3.

Table 3: **Machine Learning versus. Survey Forecasts of GDP Growth**

ML: $y_{j,t+h} = \alpha_j^{(i)} + \beta_{j\mathbb{F}}^{(i)} \mathbb{F}_t^{(i)} [y_{j,t+h}] + \mathbf{B}_{j\mathcal{Z}}^{(i)} \mathcal{Z}_{jt} + \epsilon_{jt+h}$

Survey of Professional Forecasters (SPF)

Percentile	Median	Mean	5th	10th	20th
$MSE_{\mathbb{E}}$	2.35	2.41	2.22	2.27	2.12
$MSE_{\mathbb{F}}$	2.63	2.60	3.10	2.73	2.59
$MSE_{\mathbb{E}}/MSE_{\mathbb{F}}$	0.89	0.83	0.72	0.83	0.82
	25th	30th	40th	60th	70th
$MSE_{\mathbb{E}}$	2.21	2.28	2.34	2.34	2.31
$MSE_{\mathbb{F}}$	2.57	2.57	2.60	2.70	2.81
$MSE_{\mathbb{E}}/MSE_{\mathbb{F}}$	0.86	0.89	0.90	0.87	0.82
	75th	80th	90th	95th	
$MSE_{\mathbb{E}}$	2.32	2.43	2.39	2.54	
$MSE_{\mathbb{F}}$	2.87	2.96	3.38	3.90	
$MSE_{\mathbb{E}}/MSE_{\mathbb{F}}$	0.81	0.82	0.71	0.65	

Michigan Survey of Consumers (SOC)

Percentile	Median
$MSE_{\mathbb{E}}$	2.41
$MSE_{\mathbb{F}}$	3.24
$MSE_{\mathbb{E}}/MSE_{\mathbb{F}}$	0.74

Blue Chip Financial Forecasts (BC)

Percentile	Median	Mean	5th	10th	20th
$MSE_{\mathbb{E}}$	2.14	2.29	2.29	2.08	2.37
$MSE_{\mathbb{F}}$	2.81	2.77	2.97	2.76	2.67
$MSE_{\mathbb{E}}/MSE_{\mathbb{F}}$	0.76	0.83	0.77	0.75	0.89
	25th	30th	40th	60th	70th
$MSE_{\mathbb{E}}$	2.19	2.20	2.11	2.20	2.29
$MSE_{\mathbb{F}}$	2.68	2.71	2.76	2.89	2.99
$MSE_{\mathbb{E}}/MSE_{\mathbb{F}}$	0.82	0.81	0.77	0.76	0.73
	75th	80th	90th	95th	
$MSE_{\mathbb{E}}$	2.23	2.23	2.28	2.54	
$MSE_{\mathbb{F}}$	3.06	3.18	3.50	3.82	
$MSE_{\mathbb{E}}/MSE_{\mathbb{F}}$	0.70	0.65	0.67		

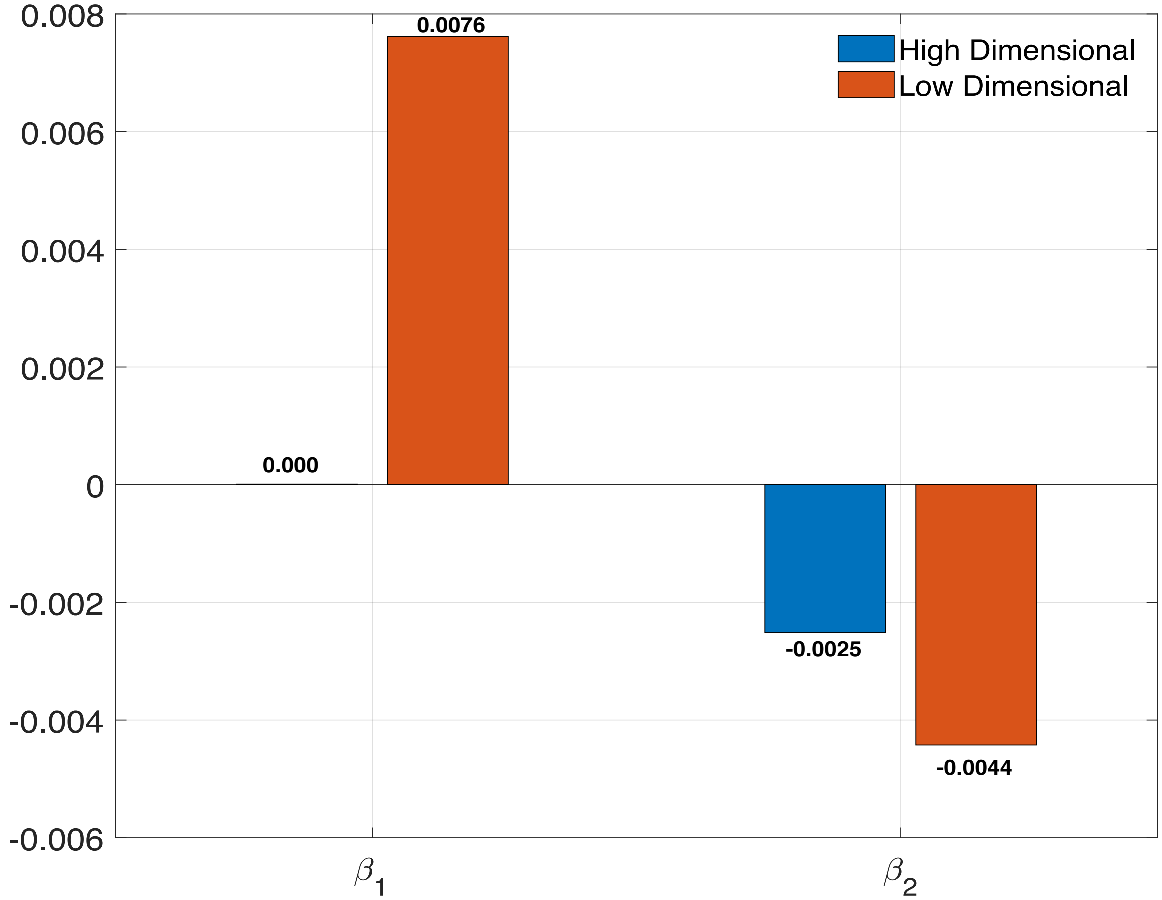
Machine v.s. survey mean-square-forecast errors for GDP growth. $MSE_{\mathbb{E}}$ and $MSE_{\mathbb{F}}$ denote the machine learning and survey mean-squared-forecast-errors, respectively, computed for 4-quarter-ahead forecasts and averaged over the evaluation period. The evaluation period for the Survey of Professional Forecasters (SPF) is 1995:Q1 to 2018:Q2; for the Michigan Survey of Consumers (SOC) is 1995:Q1 to 2018:Q2; and for the Bluechip (BC)survey is 1997:Q1 to 2018:Q2. The full estimation sample spans the periods 1969:Q3 to 2018:Q3.

Table 4. **Decomposition of Variance**

	Fraction Variation in Real GDP (%)				Fraction Variation in Real Investment (%)			
Explained by:	$\overline{bias}_t^{\pi}$	FFR	$\overline{bias}_t^{\Delta y}$	FFR	$\overline{bias}_t^{\pi}$	FFR	$\overline{bias}_t^{\Delta y}$	FFR
$k = 3$	1.21	0.10	0.45	0.04	4.68	1.42	0.62	1.60
$k = 12$	1.56	1.44	5.18	1.96	15.36	13.61	3.93	10.27
$k = \infty$	1.99	6.16	8.91	5.69	13.97	18.97	5.70	11.67
$\max k$	25	32	34	32	11	∞	∞	163
$k = \max$	3.94	9.04	9.97	6.95	15.39	18.97	5.70	11.67
	Fraction Variation in GDP Deflator (%):				Fraction Variation in S&P 500 (%)			
Explained by:	$\overline{bias}_t^{\pi}$	FFR	$\overline{bias}_t^{\Delta y}$	FFR	$\overline{bias}_t^{\pi}$	FFR	$\overline{bias}_t^{\Delta y}$	FFR
$k = 3$	1.94	0.70	0.13	0.60	0.03	1.03	1.39	0.64
$k = 12$	6.50	2.09	0.29	1.18	2.45	1.51	2.61	1.04
$k = \infty$	0.79	3.90	7.66	5.22	2.73	2.03	3.17	1.22
$\max k$	10	55	125	117	52	55	23	7
$k = \max$	6.64	4.36	7.91	5.54	3.21	2.21	3.18	1.42
	Fraction of Variation in Real Wage (%):				Fraction Variation in FFR (%)			
Explained by:	$\overline{bias}_t^{\pi}$	FFR	$\overline{bias}_t^{\Delta y}$	FFR	$\overline{bias}_t^{\pi}$	FFR	$\overline{bias}_t^{\Delta y}$	FFR
$k = 3$	2.92	0.16	0.22	0.21	0.12	70.80	3.25	66.55
$k = 12$	3.49	2.64	4.94	5.04	0.10	44.93	3.27	42.83
$k = \infty$	3.68	5.48	6.63	6.48	0.73	34.41	2.89	32.62
$\max k$	6	25	25	23	71	1	6	1
$k = \max$	5.02	6.13	7.34	7.59	0.73	75.93	4.26	72.54

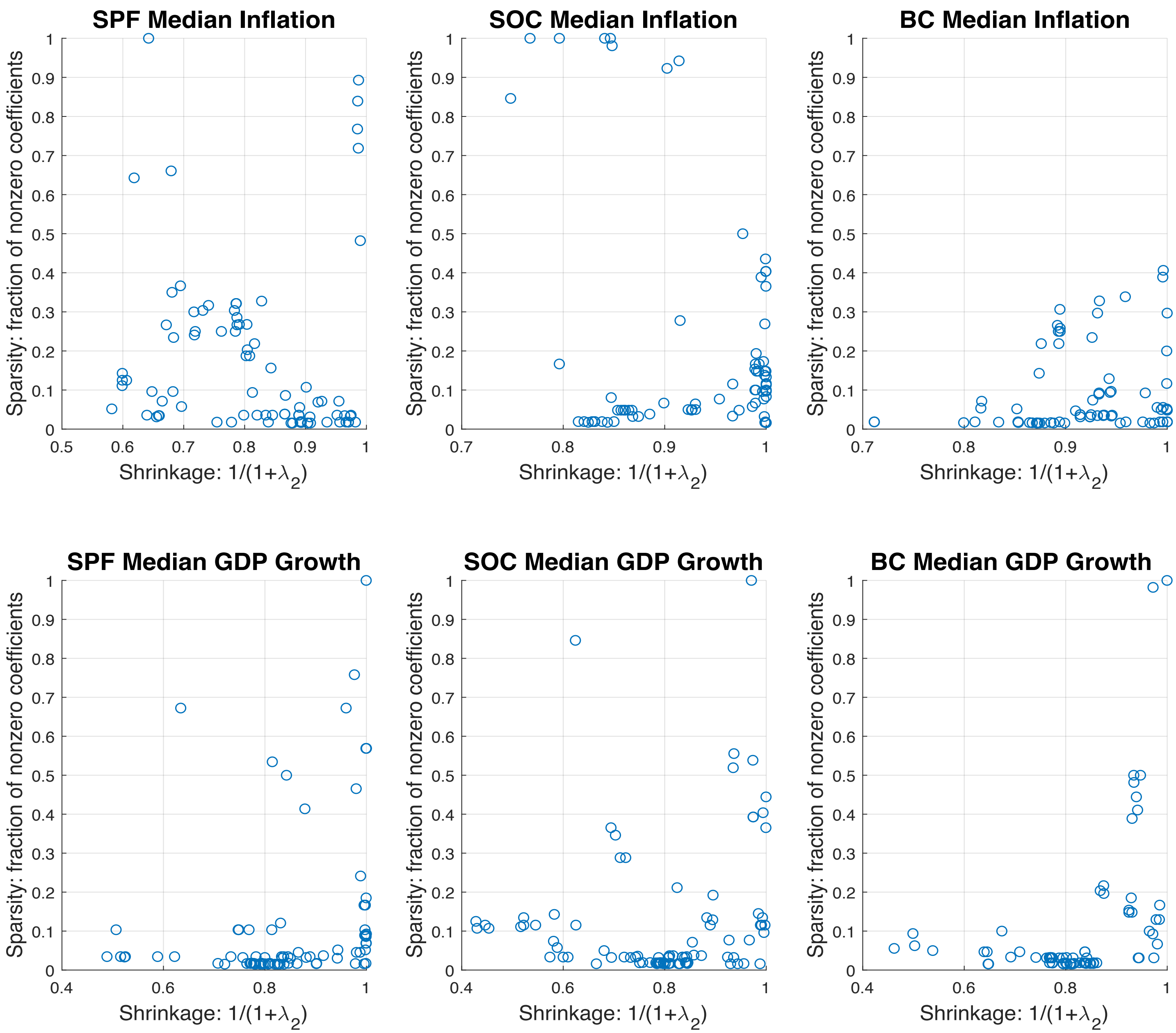
Forecast error variance decomposition. Forecast error variances are computed from a VAR using a Cholesky factorization with the following variables in the order: log(real GDP), log(GDP deflator), log(real wage), log(real investment), log(S&P 500 Index), federal funds rate (FFR), and $\overline{bias}_t$, where $\overline{bias}_t$ is either the inflation bias index $\overline{bias}_t^{\pi}$ or the GDP growth bias index $\overline{bias}_t^{\Delta y}$. Each panel shows the fraction of forecast-error variance of the variable named in the panel title at VAR forecast horizon k that is explained by $\overline{bias}_t$ or the FFR for that VAR. The row denoted "max k" gives the horizon k for which the variable named in the column explains the maximum fraction of forecast error variance. The row denoted "$k = \max$" gives the fraction of forecast error variance explained at max k. The data are quarterly and span the period 1995:Q1 -2018:Q2.

Figure 1: High- v.s. Low-dimensional Out-of-sample Forecasts



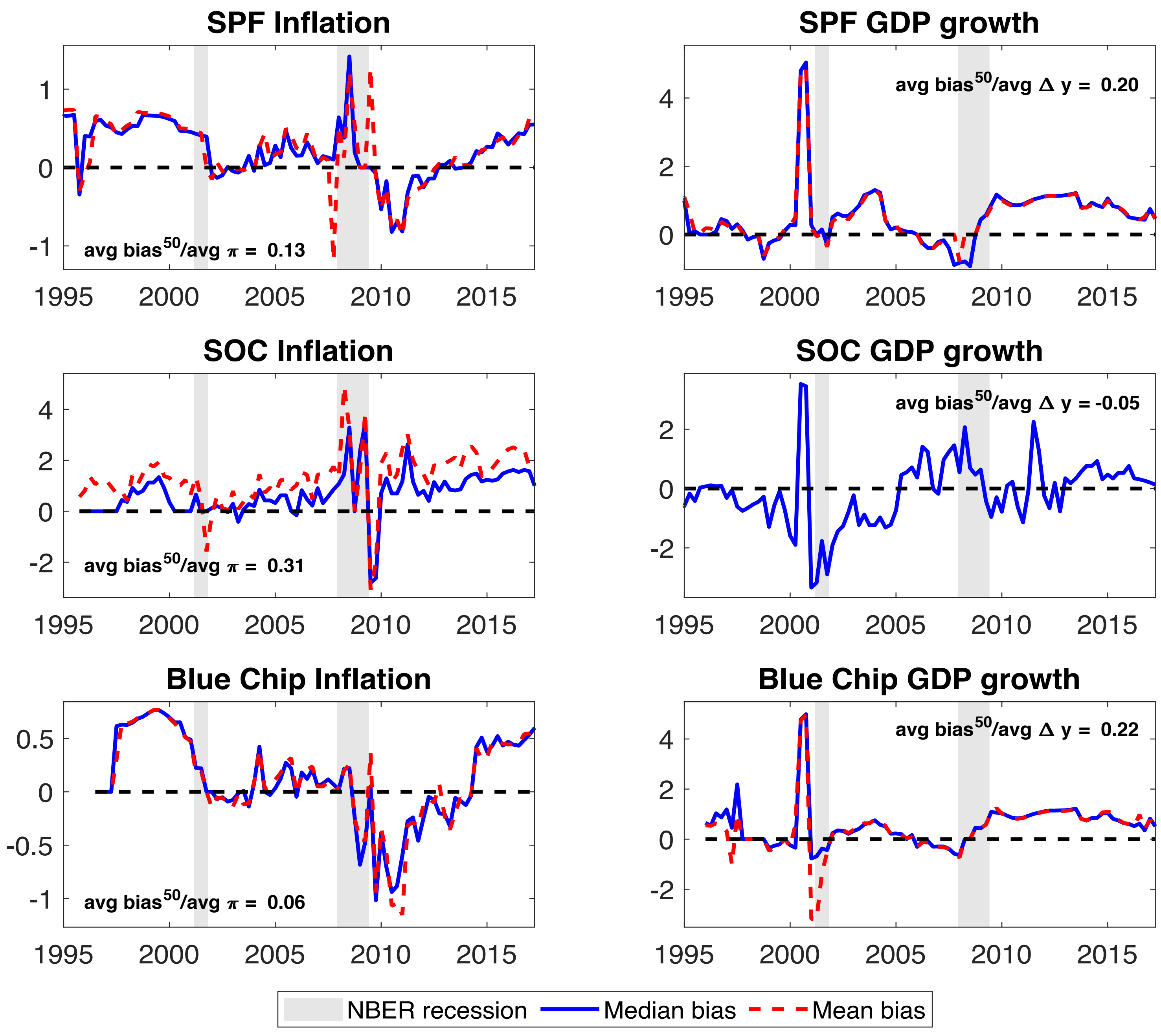
Autoregressive Coefficients in high- v.s. low-dimensional out-of-sample forecasts. Average autoregressive coefficients from one-year-ahead rolling regressions of real GDP growth on predictors. $\boldsymbol{\beta}_1$ is the average coefficient on the first AR lag; $\boldsymbol{\beta}_2$ is the average coefficient on the second. The high dimension estimation entertains very large numbers of potential predictors, in addition to the autoregressive lags, while the low dimension setting uses only two additional predictors. The sample spans 1995:Q1-2018:Q2.

Figure 2: Degree of Sparsity and Shrinkage



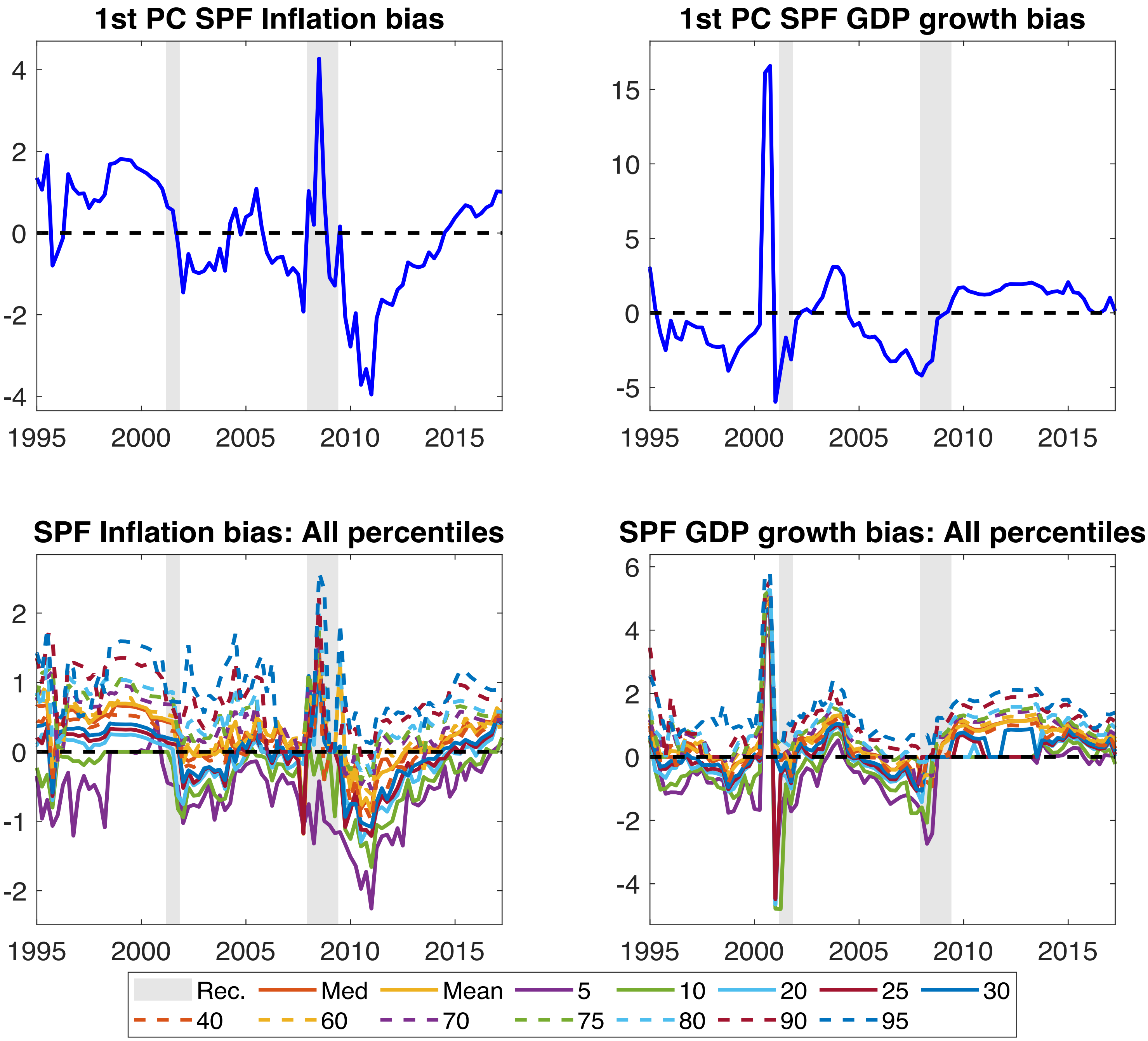
Degree of Sparsity and Shrinkage. The figure displays a scatterplot of the strength of the ridge and LASSO penalties estimated from training samples over time for predicting median inflation or real GDP growth. For each observation in the evaluation sample from 1995:1-2018:Q2 (94 observations), the y-axis displays the degree of sparsity implied by the estimated L^1 penalty, λ_1, in units of the fraction of non-zero regression coefficients, and the x-axis displays the degree of shrinkage implied by the estimated L^2 penalty, λ_2 in units of $1/(1+\lambda_2)$.

Figure 3: Biases in the Mean and Median Survey Forecasts



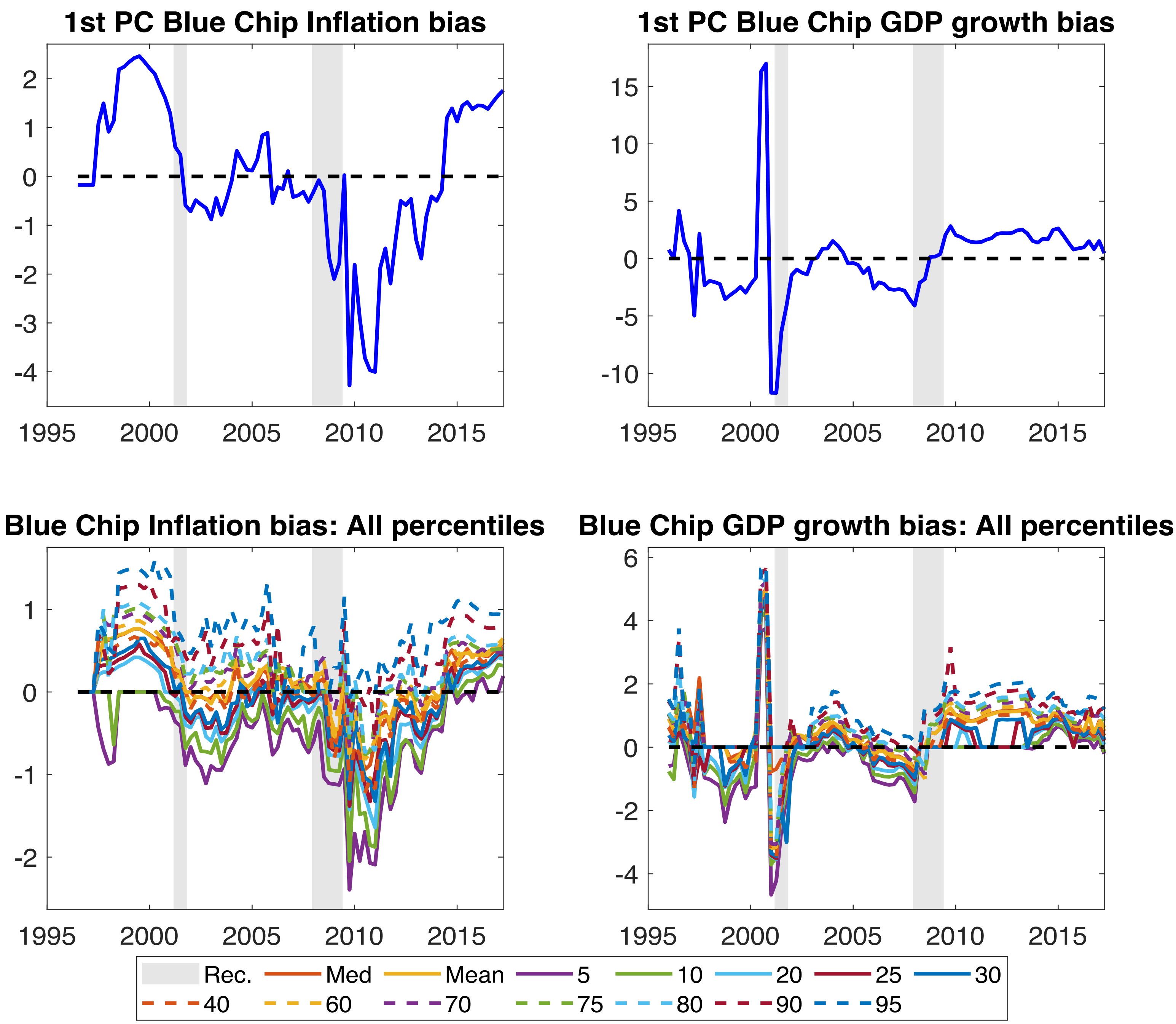
Biases in the consensus forecasts. The figure reports the time series $bias_{j,t+h}^{(i)} = \mathbb{F}_t^{(i)}[y_{j,t+h}] - \mathbb{E}_t^{(i)}[y_{j,t+h}]$ for $i = 50, mean$. NBER recessions are shown with grey shaded bars. The sample spans the period 1995:Q1-2018:Q2.

Figure 4: Common and Heterogeneous Distortions in the SPF



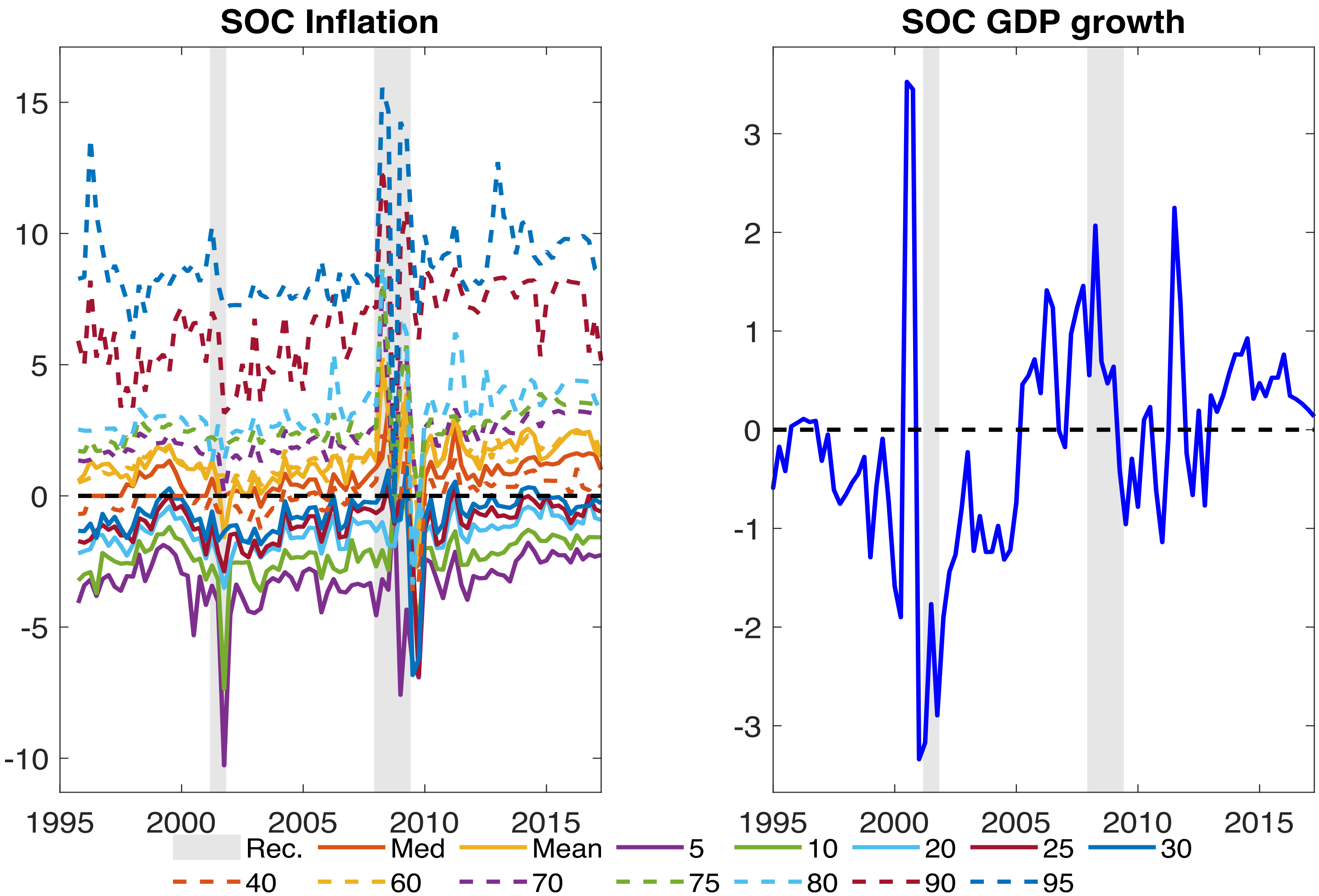
Biases in the SCF. The figure reports the time series $bias_{j,t+h}^{(i)} = \mathbb{F}_t^{(i)}\left[y_{j,t+h}\right] - \mathbb{E}_t^{(i)}\left[y_{j,t+h}\right]$. NBER recessions are shown with grey shaded bars. The sample is 1995:Q1-2018:Q2.

Figure 5: Common and Heterogeneous Distortions in the Blue Chip



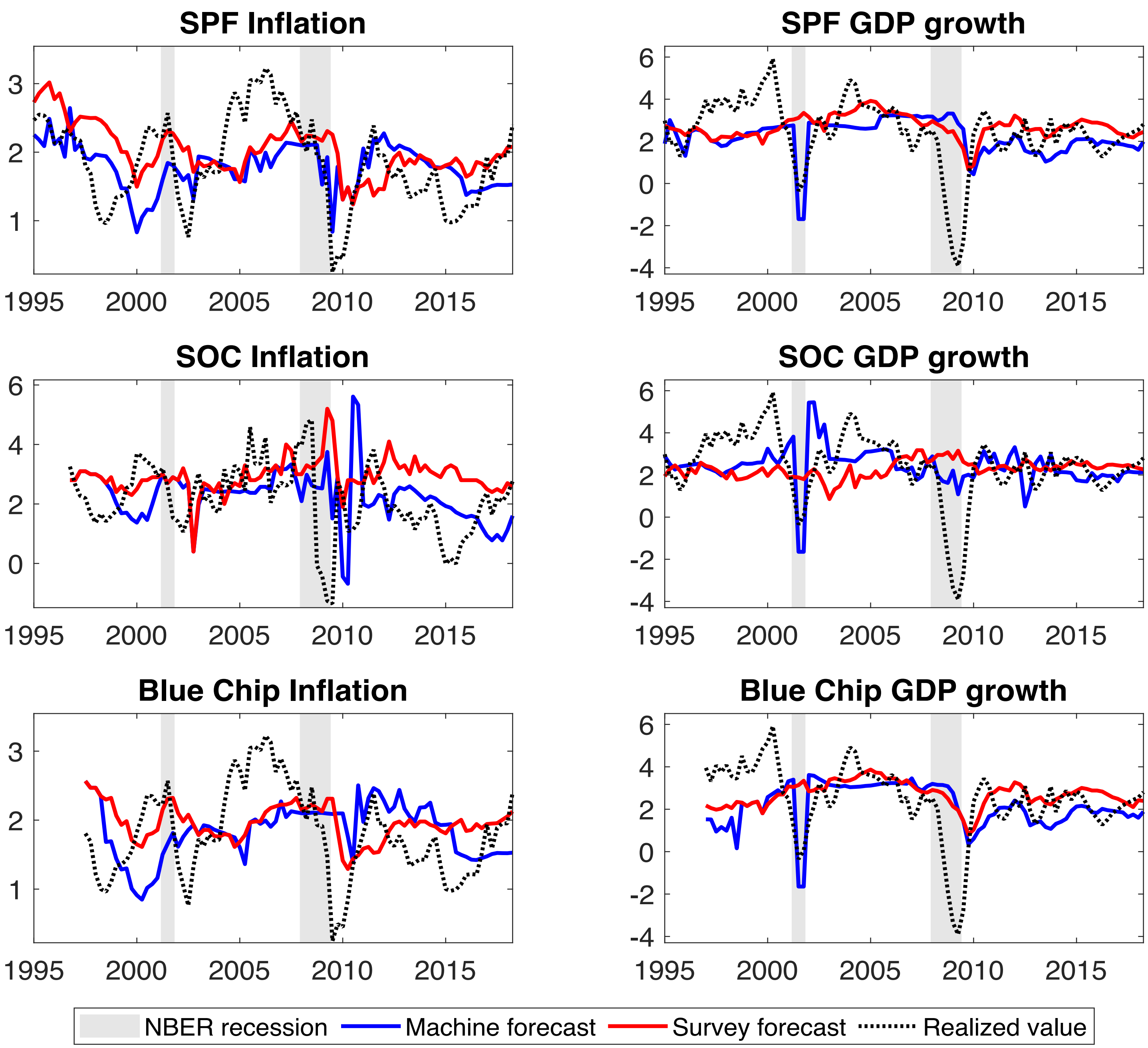
Biases in the Blue Chip. The figure reports the time series $bias_{j,t+h}^{(i)} = \mathbb{F}_t^{(i)}[y_{j,t+h}] - \mathbb{E}_t^{(i)}[y_{j,t+h}]$. NBER recessions are shown with grey shaded bars. The sample is 1997:Q1-2017:Q1.

Figure 6: Common and Heterogeneous Distortions in the SOC



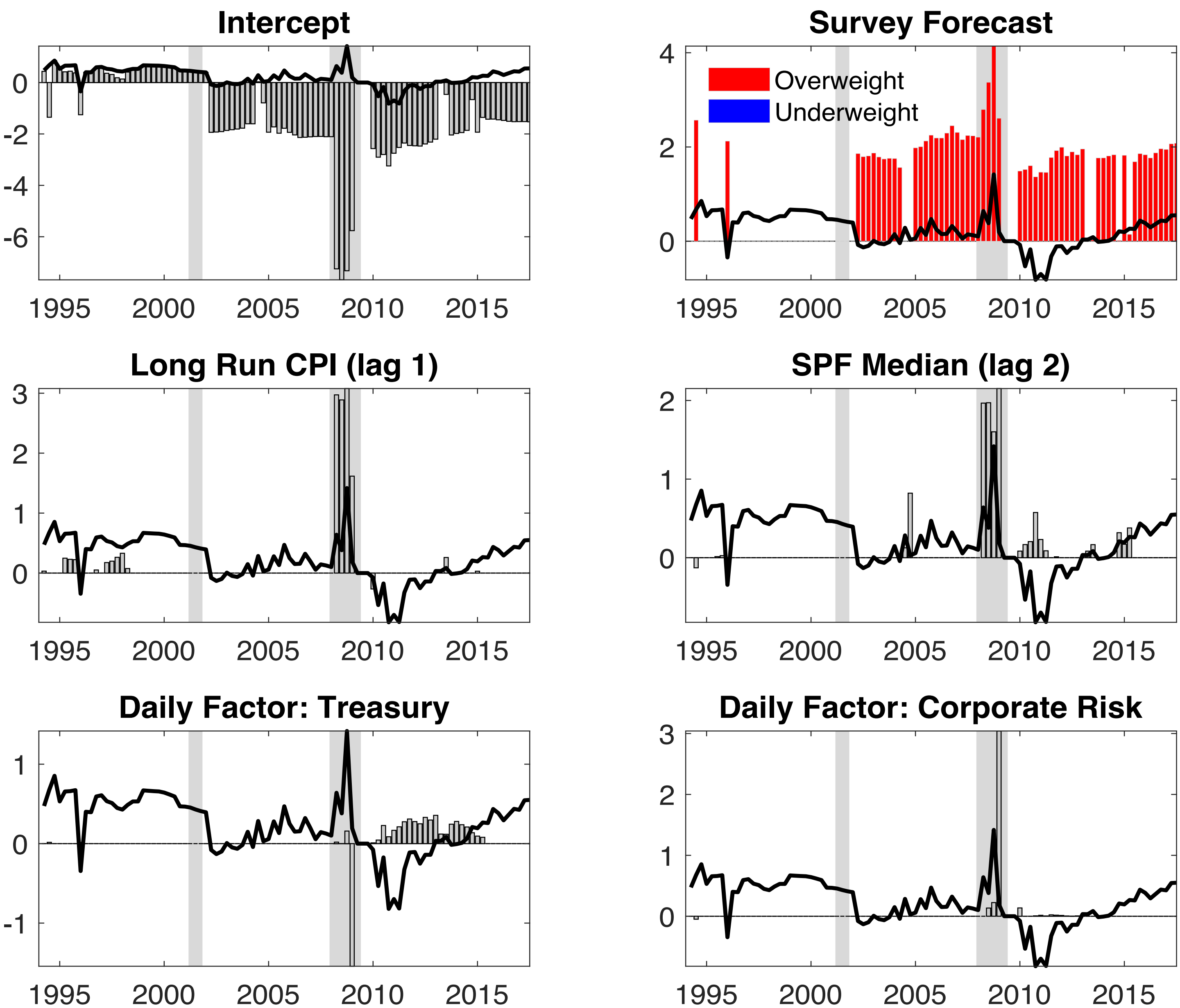
Biases in the SOC. The figure reports the time series $bias^{(i)}_{j,t+h} = \mathbb{F}^{(i)}_t [y_{j,t+h}] - \mathbb{E}^{(i)}_t [y_{j,t+h}]$. NBER recessions are shown with grey shaded bars. The sample is 1995:Q1-2018:Q2.

Figure 7: Forecasted versus Actual Inflation, GDP Growth



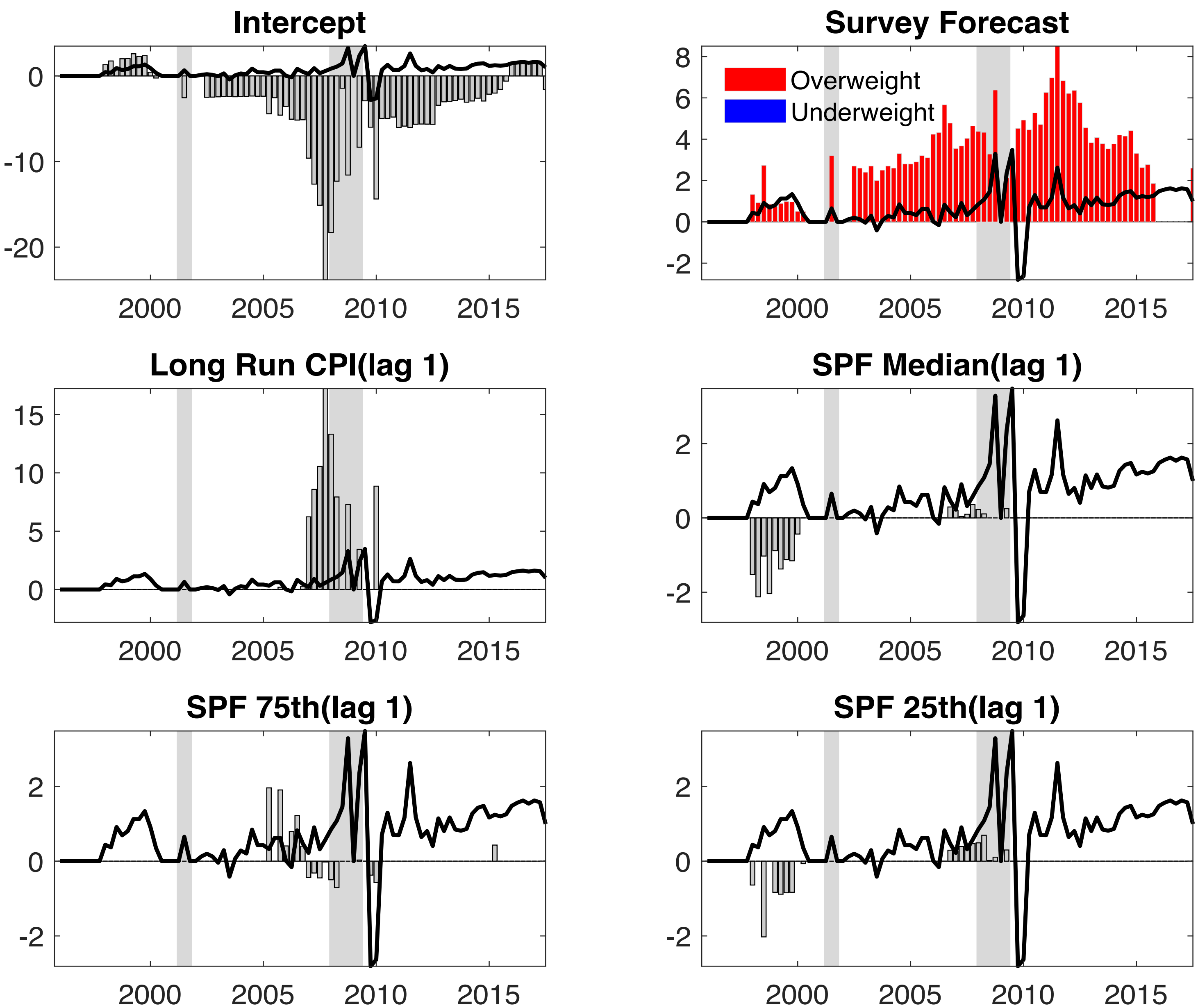
Forecasted and Actual variables. For each variable and survey, the figure reports the median survey forecast of inflation or GDP growth over the next 4 quarters, the corresponding the machine forecast, and the actual inflation or GDP growth during this period. NBER recessions are shown with grey shaded bars. The sample is 1995:Q1-2018:Q2.

Figure 8: Bias Decomposition: SPF Inflation Median Forecast



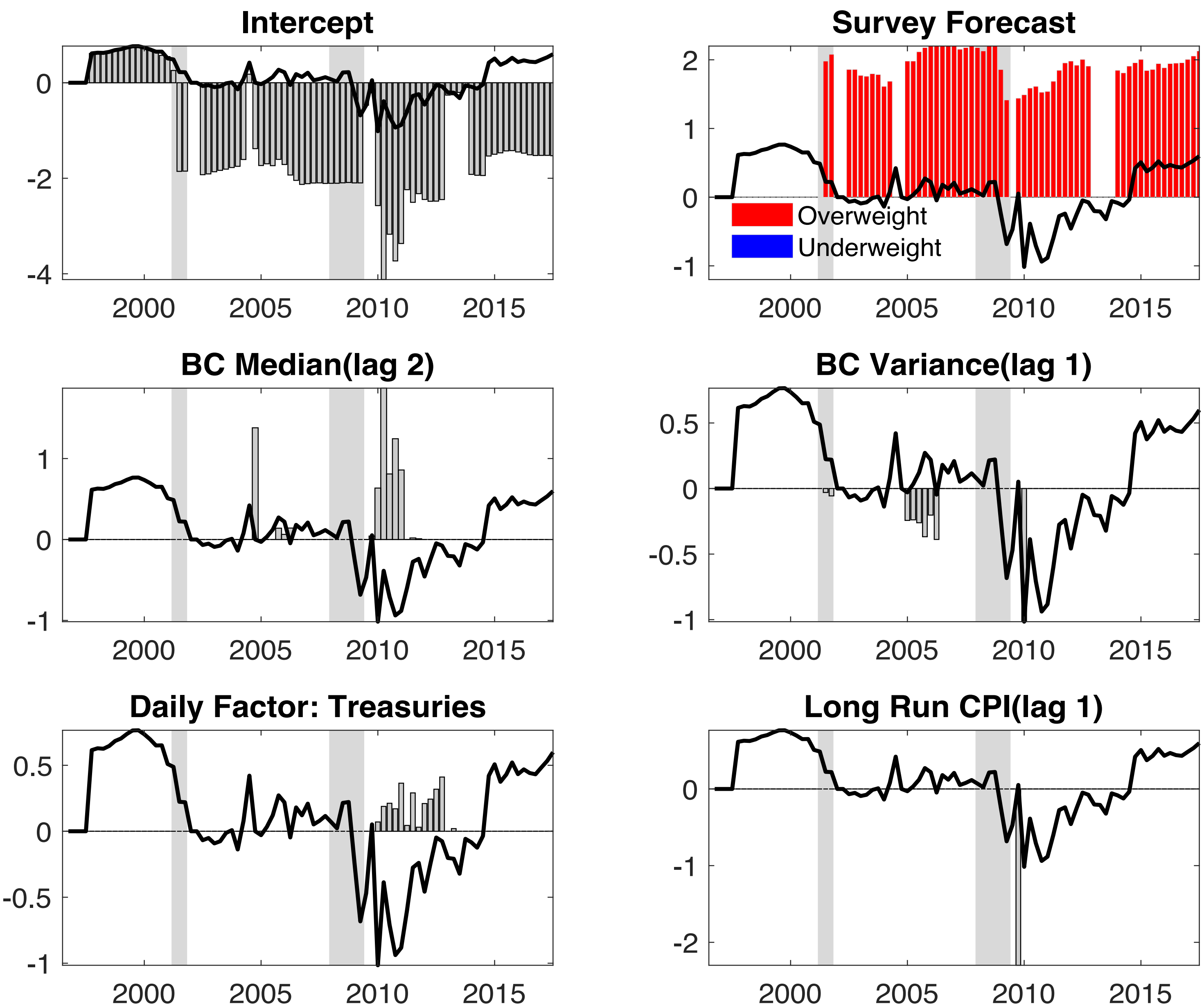
Decomposition of Bias. The figure plots contributors to the median bias $\mathbb{F}_t^{(50)}[y_{j,t+h}] - \mathbb{E}_t^{(50)}[y_{j,t+h}] = -\hat{\alpha}_j^{(50)} + \left(1-\hat{\beta}_{j\mathbb{F}}^{(50)}\right)\mathbb{F}_t^{(50)}[y_{j,t+h}] - \hat{B}_{j\mathcal{Z}}^{(50)\prime}\mathcal{Z}_{jt}$ at each time t. The solid black lines in each subpanel plot the median overall bias, $F_t^{(50)}[y_{j,t+h}] - E_t^{(50)}[y_{j,t+h}]$. The barchart in the "Intercept" subpanel reports $-\hat{\alpha}_j^{(50)}$; the barchart in the "Survey Forecast" panel reports $\left(1-\hat{\beta}_{j\mathbb{F}}^{(50)}\right)\mathbb{F}_t^{(50)}[y_{j,t+h}]$. The barcharts in the remaining subpanels report $-\hat{B}_{j\mathcal{Z}}^{(50)\prime}\mathcal{Z}_{jt}$ for the top four most important predictor contributors to the bias, as measured by the absolute sum of contributions over the evaluation sample. Red bars indicate that the survey forecast was given too much weight relative to the machine efficient forecast, corresponding to $\left(1-\hat{\beta}_{j\mathbb{F}}^{(50)}\right) > 0$. Blue bars indicate that the survey forecast was given too little weight relative to the machine efficient forecast, corresponding to $\left(1-\hat{\beta}_{j\mathbb{F}}^{(50)}\right) < 0$. NBER recessions are shown with grey shaded bars. The evaluation sample spans the period 1995:Q1-2018:Q2.

Figure 9: Bias Decomposition: SOC Inflation Median Forecast



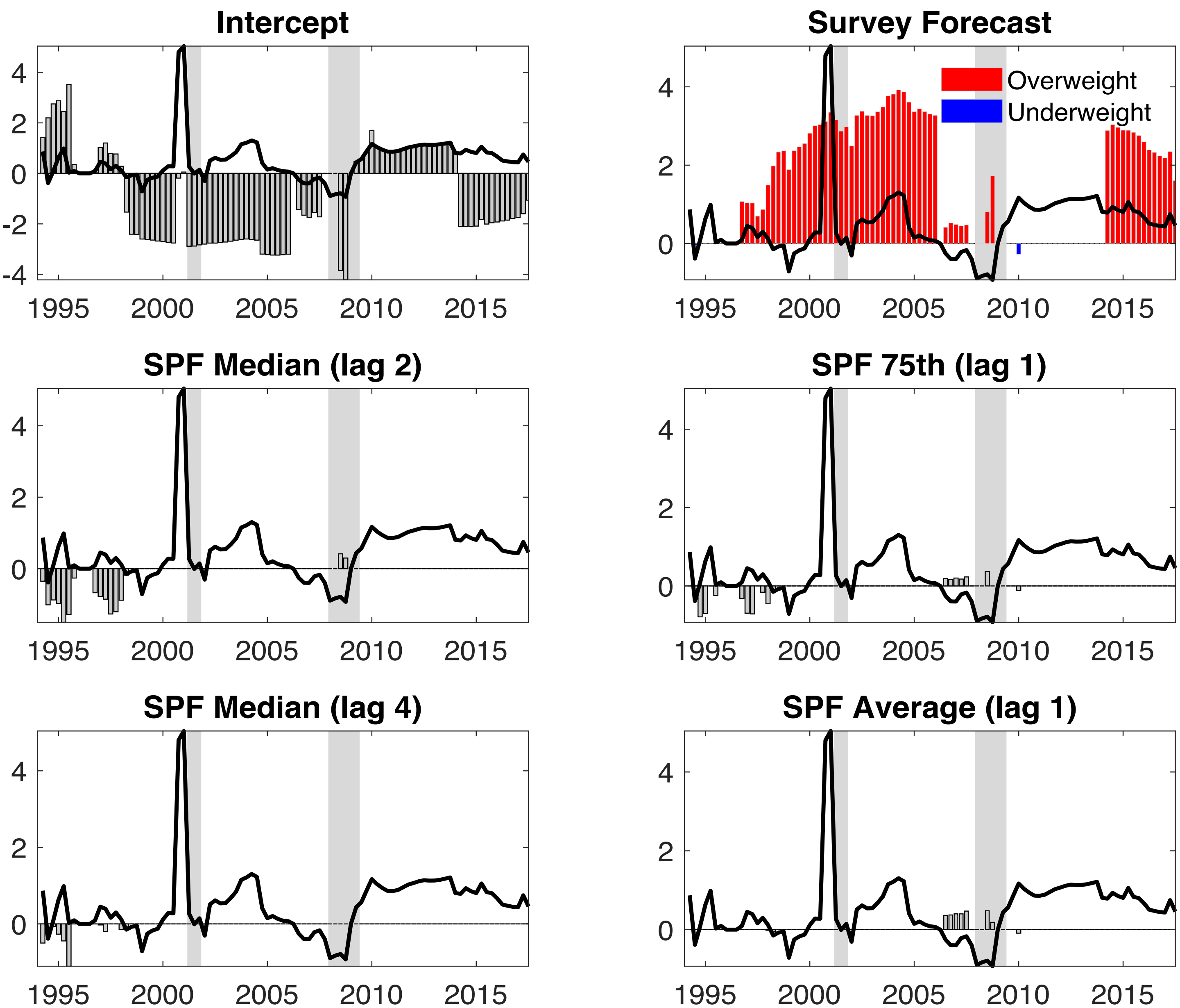
Decomposition of Bias. The figure plots contributors to the median bias $\mathbb{F}_t^{(50)}[y_{j,t+h}] - \mathbb{E}_t^{(50)}[y_{j,t+h}] = -\hat{\alpha}_j^{(50)} + \left(1 - \hat{\beta}_{j\mathbb{F}}^{(50)}\right)\mathbb{F}_t^{(50)}[y_{j,t+h}] - \hat{B}_{j\mathcal{Z}}^{(50)\prime}\mathcal{Z}_{jt}$ at each time t. The solid black lines in each subpanel plot the median overall bias, $F_t^{(50)}[y_{j,t+h}] - E_t^{(50)}[y_{j,t+h}]$. The barchart in the "Intercept" subpanel reports $-\hat{\alpha}_j^{(50)}$; the barchart in the "Survey Forecast" panel reports $\left(1 - \hat{\beta}_{j\mathbb{F}}^{(50)}\right)\mathbb{F}_t^{(50)}[y_{j,t+h}]$. The barcharts in the remaining subpanels report $-\hat{B}_{j\mathcal{Z}}^{(50)\prime}\mathcal{Z}_{jt}$ for the top four most important predictor contributors to the bias, as measured by the absolute sum of contributions over the evaluation sample. Red bars indicate that the survey forecast was given too much weight relative to the machine efficient forecast, corresponding to $\left(1 - \hat{\beta}_{j\mathbb{F}}^{(50)}\right) > 0$. Blue bars indicate that the survey forecast was given too little weight relative to the machine efficient forecast, corresponding to $\left(1 - \hat{\beta}_{j\mathbb{F}}^{(50)}\right) < 0$. NBER recessions are shown with grey shaded bars. The evaluation sample spans the period 1995:Q1-2018:Q2.

Figure 10: Bias Decomposition: BC Inflation Median Forecast



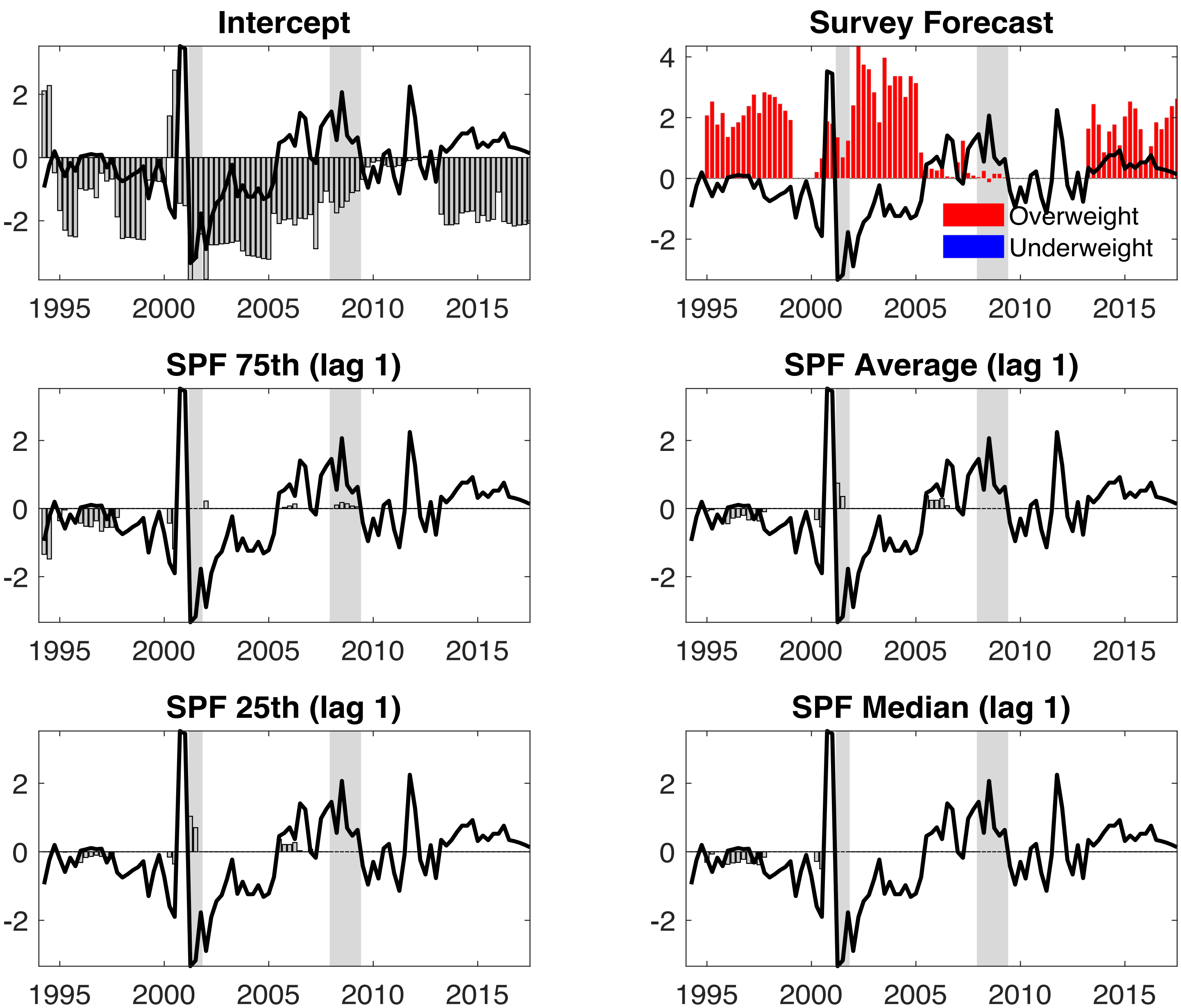
Decomposition of Bias. The figure plots contributors to the median bias $\mathbb{F}_t^{(50)}[y_{j,t+h}] - \mathbb{E}_t^{(50)}[y_{j,t+h}] = -\hat{\alpha}_j^{(50)} + \left(1 - \hat{\beta}_{j\mathbb{F}}^{(50)}\right)\mathbb{F}_t^{(50)}[y_{j,t+h}] - \hat{B}_{j\mathcal{Z}}^{(50)\prime}\mathcal{Z}_{jt}$ at each time t. The solid black lines in each subpanel plot the median overall bias, $F_t^{(50)}[y_{j,t+h}] - E_t^{(50)}[y_{j,t+h}]$. The barchart in the "Intercept" subpanel reports $-\hat{\alpha}_j^{(50)}$; the barchart in the "Survey Forecast" panel reports $\left(1 - \hat{\beta}_{j\mathbb{F}}^{(50)}\right)\mathbb{F}_t^{(50)}[y_{j,t+h}]$. The barcharts in the remaining subpanels report $-\hat{B}_{j\mathcal{Z}}^{(50)\prime}\mathcal{Z}_{jt}$ for the top four most important predictor contributors to the bias, as measured by the absolute sum of contributions over the evaluation sample. Red bars indicate that the survey forecast was given too much weight relative to the machine efficient forecast, corresponding to $\left(1 - \hat{\beta}_{j\mathbb{F}}^{(50)}\right) > 0$. Blue bars indicate that the survey forecast was given too little weight relative to the machine efficient forecast, corresponding to $\left(1 - \hat{\beta}_{j\mathbb{F}}^{(50)}\right) < 0$. NBER recessions are shown with grey shaded bars. The evaluation sample spans the period 1995:Q1-2018:Q2.

Figure 11: Bias Decomposition: SPF Real GDP Growth Median Forecast



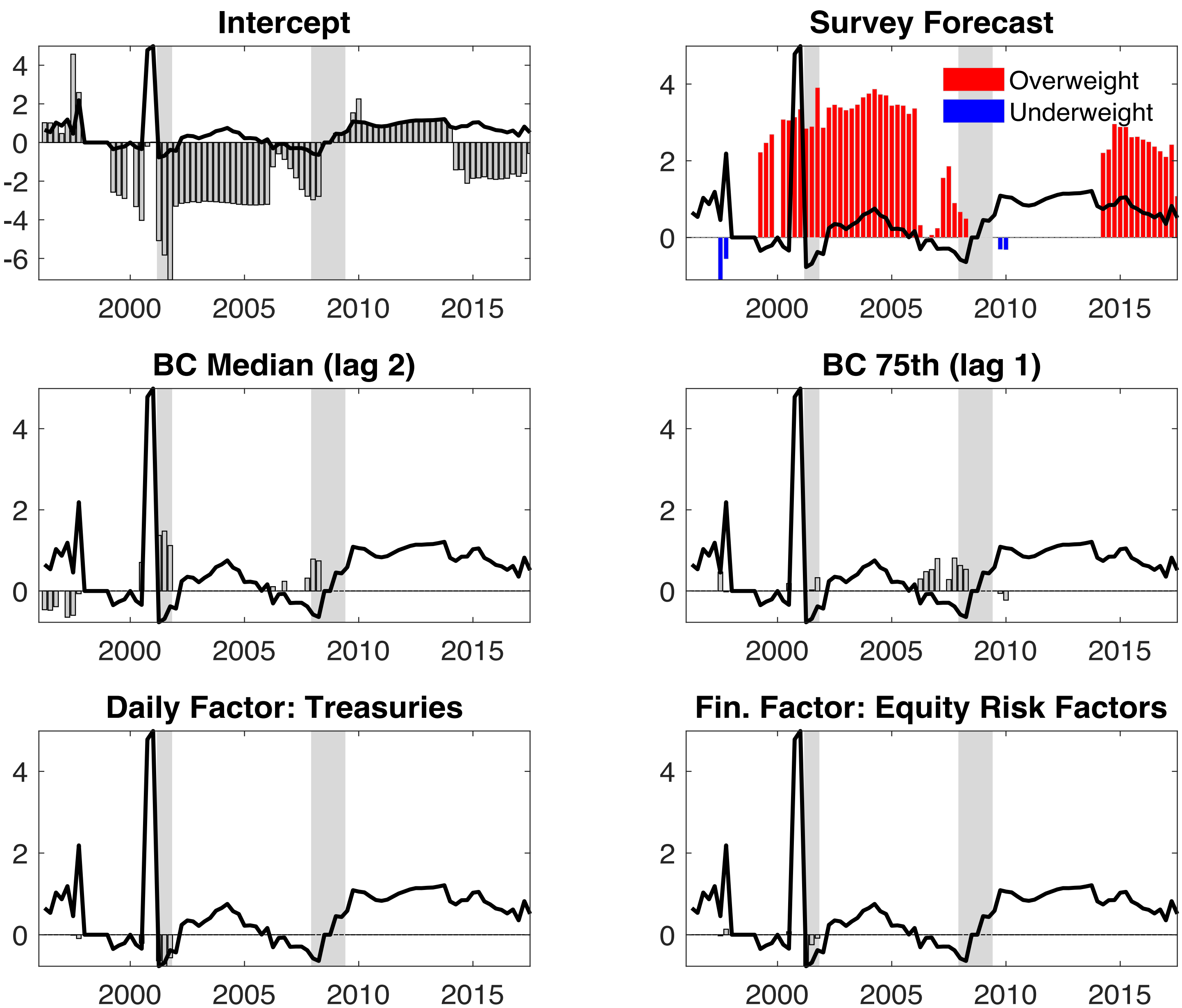
Decomposition of Bias. The figure plots contributors to the median bias $\mathbb{F}_t^{(50)}[y_{j,t+h}] - \mathbb{E}_t^{(50)}[y_{j,t+h}] = -\hat{\alpha}_j^{(50)} + \left(1 - \hat{\beta}_{j\mathbb{F}}^{(50)}\right)\mathbb{F}_t^{(50)}[y_{j,t+h}] - \hat{B}_{j\mathcal{Z}}^{(50)\prime}\mathcal{Z}_{jt}$ at each time t. The solid black lines in each subpanel plot the median overall bias, $F_t^{(50)}[y_{j,t+h}] - E_t^{(50)}[y_{j,t+h}]$. The barchart in the "Intercept" subpanel reports $-\hat{\alpha}_j^{(50)}$; the barchart in the "Survey Forecast" panel reports $\left(1 - \hat{\beta}_{j\mathbb{F}}^{(50)}\right)\mathbb{F}_t^{(50)}[y_{j,t+h}]$. The barcharts in the remaining subpanels report $-\hat{B}_{j\mathcal{Z}}^{(50)\prime}\mathcal{Z}_{jt}$ for the top four most important predictor contributors to the bias, as measured by the absolute sum of contributions over the evaluation sample. Red bars indicate that the survey forecast was given too much weight relative to the machine efficient forecast, corresponding to $\left(1 - \hat{\beta}_{j\mathbb{F}}^{(50)}\right) > 0$. Blue bars indicate that the survey forecast was given too little weight relative to the machine efficient forecast, corresponding to $\left(1 - \hat{\beta}_{j\mathbb{F}}^{(50)}\right) < 0$. NBER recessions are shown with grey shaded bars. The evaluation sample spans the period 1995:Q1-2018:Q2.

Figure 12: Bias Decomposition: SOC Real GDP Growth Median Forecast



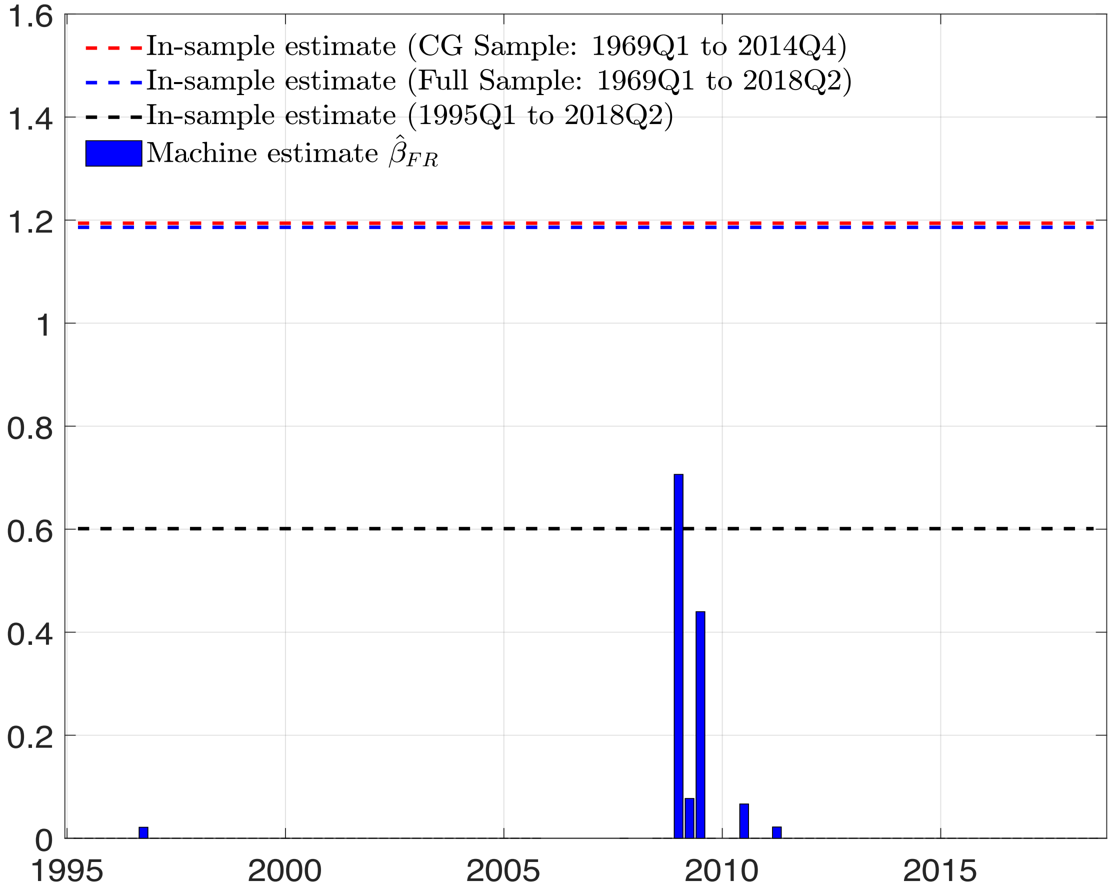
Decomposition of Bias. The figure plots contributors to the median bias $\mathbb{F}_t^{(50)}[y_{j,t+h}] - \mathbb{E}_t^{(50)}[y_{j,t+h}] = -\hat{\alpha}_j^{(50)} + \left(1 - \hat{\beta}_{j\mathbb{F}}^{(50)}\right)\mathbb{F}_t^{(50)}[y_{j,t+h}] - \hat{B}_{j\mathcal{Z}}^{(50)\prime}\mathcal{Z}_{jt}$ at each time t. The solid black lines in each subpanel plot the median overall bias, $F_t^{(50)}[y_{j,t+h}] - E_t^{(50)}[y_{j,t+h}]$. The barchart in the "Intercept" subpanel reports $-\hat{\alpha}_j^{(50)}$; the barchart in the "Survey Forecast" panel reports $\left(1 - \hat{\beta}_{j\mathbb{F}}^{(50)}\right)\mathbb{F}_t^{(50)}[y_{j,t+h}]$. The barcharts in the remaining subpanels report $-\hat{B}_{j\mathcal{Z}}^{(50)\prime}\mathcal{Z}_{jt}$ for the top four most important predictor contributors to the bias, as measured by the absolute sum of contributions over the evaluation sample. Red bars indicate that the survey forecast was given too much weight relative to the machine efficient forecast, corresponding to $\left(1 - \hat{\beta}_{j\mathbb{F}}^{(50)}\right) > 0$. Blue bars indicate that the survey forecast was given too little weight relative to the machine efficient forecast, corresponding to $\left(1 - \hat{\beta}_{j\mathbb{F}}^{(50)}\right) < 0$. NBER recessions are shown with grey shaded bars. The evaluation sample spans the period 1995:Q1-2018:Q2.

Figure 13: Bias Decomposition: BC Real GDP Growth Median Forecast



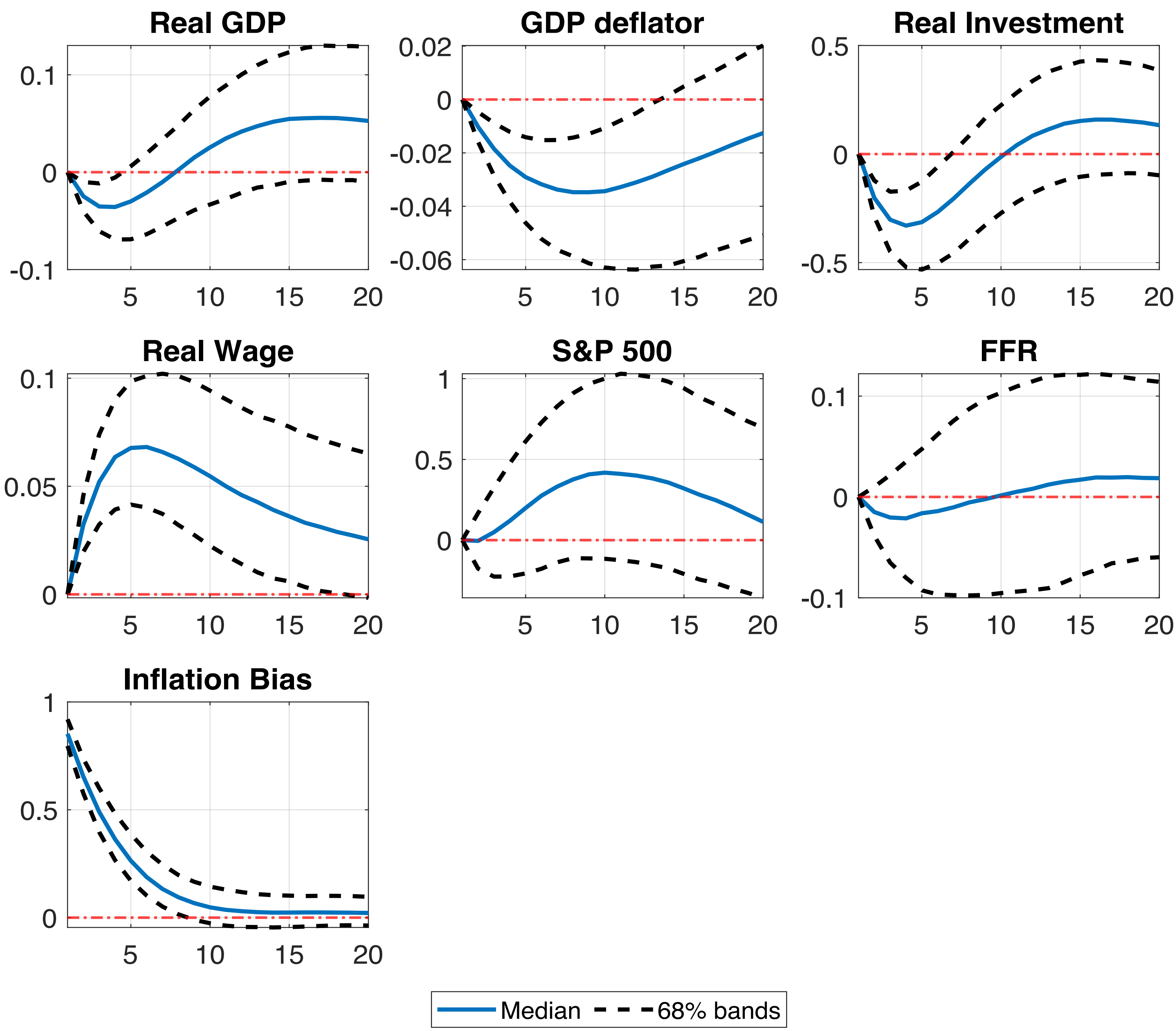
Decomposition of Bias. The figure plots contributors to the median bias $\mathbb{F}_t^{(50)}[y_{j,t+h}] - \mathbb{E}_t^{(50)}[y_{j,t+h}] = -\hat{\alpha}_j^{(50)} + \left(1 - \hat{\beta}_{j\mathbb{F}}^{(50)}\right)\mathbb{F}_t^{(50)}[y_{j,t+h}] - \hat{B}_{j\mathcal{Z}}^{(50)\prime}\mathcal{Z}_{jt}$ at each time t. The solid black lines in each subpanel plot the median overall bias, $F_t^{(50)}[y_{j,t+h}] - E_t^{(50)}[y_{j,t+h}]$. The barchart in the "Intercept" subpanel reports $-\hat{\alpha}_j^{(50)}$; the barchart in the "Survey Forecast" panel reports $\left(1 - \hat{\beta}_{j\mathbb{F}}^{(50)}\right)\mathbb{F}_t^{(50)}[y_{j,t+h}]$. The barcharts in the remaining subpanels report $-\hat{B}_{j\mathcal{Z}}^{(50)\prime}\mathcal{Z}_{jt}$ for the top four most important predictor contributors to the bias, as measured by the absolute sum of contributions over the evaluation sample. Red bars indicate that the survey forecast was given too much weight relative to the machine efficient forecast, corresponding to $\left(1 - \hat{\beta}_{j\mathbb{F}}^{(50)}\right) > 0$. Blue bars indicate that the survey forecast was given too little weight relative to the machine efficient forecast, corresponding to $\left(1 - \hat{\beta}_{j\mathbb{F}}^{(50)}\right) < 0$. NBER recessions are shown with grey shaded bars. The evaluation sample spans the period 1995:Q1-2018:Q2.

Figure 14: Coefficient on Forecast Revisions



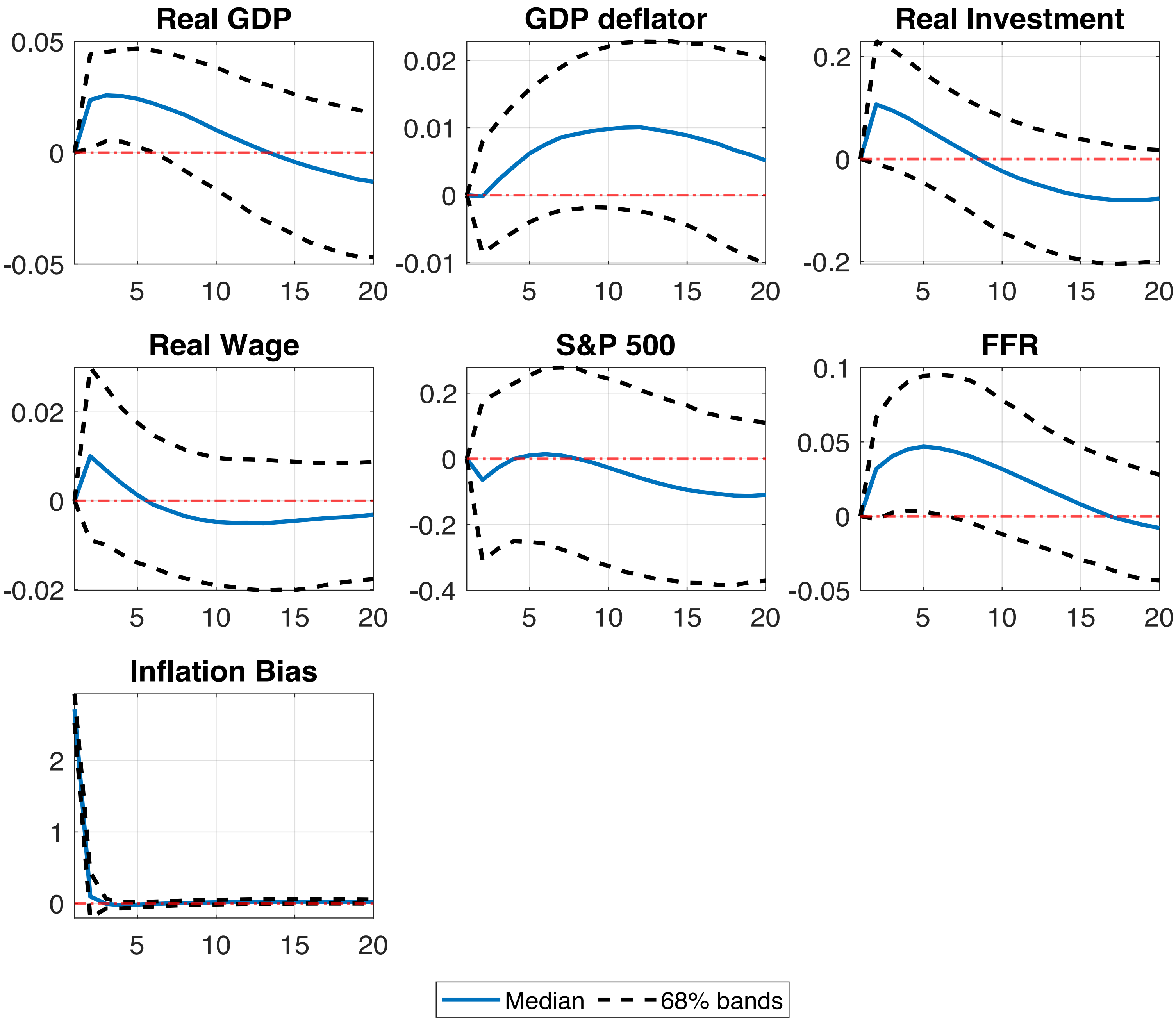
Coefficient on Forecast Revisions. The blue bar plots the estimated coefficient of from regressions of forecast errors on forecast revisions for the mean of the SPF inflation forecast $\underbrace{\pi_{t+3} - F_t^{(\mu)}[\pi_{t+3}]}_{\text{Forecast Error}} = \alpha_j^{(\mu)} + \beta_{j\text{FR}}^{(\mu)} \left(\underbrace{F_t^{(\mu)}[\pi_{t+3}] - F_{t-1}^{(\mu)}[\pi_{t+3}]}_{\text{Forecast Revisions}} \right) + B_{jZ}^{(\mu)\prime} Z_{jt} + \epsilon_{jt+h}$. This regression sample spans the period 1995:Q1 to 2018:Q2. The dashed red line shows the estimated in-sample coefficient over CG sample 1969:Q1 to 2014:Q4. The dashed blue line shows the estimated in-sample coefficient over the full sample 1969:Q1 to 2018:Q2. The dashed black line shows the estimated in-sample coefficient over the evaluation sample used for the machine estimates 1995:Q1 to 2018:Q2.

Figure 15: Responses to Inflation Bias Shock: SPF and BC



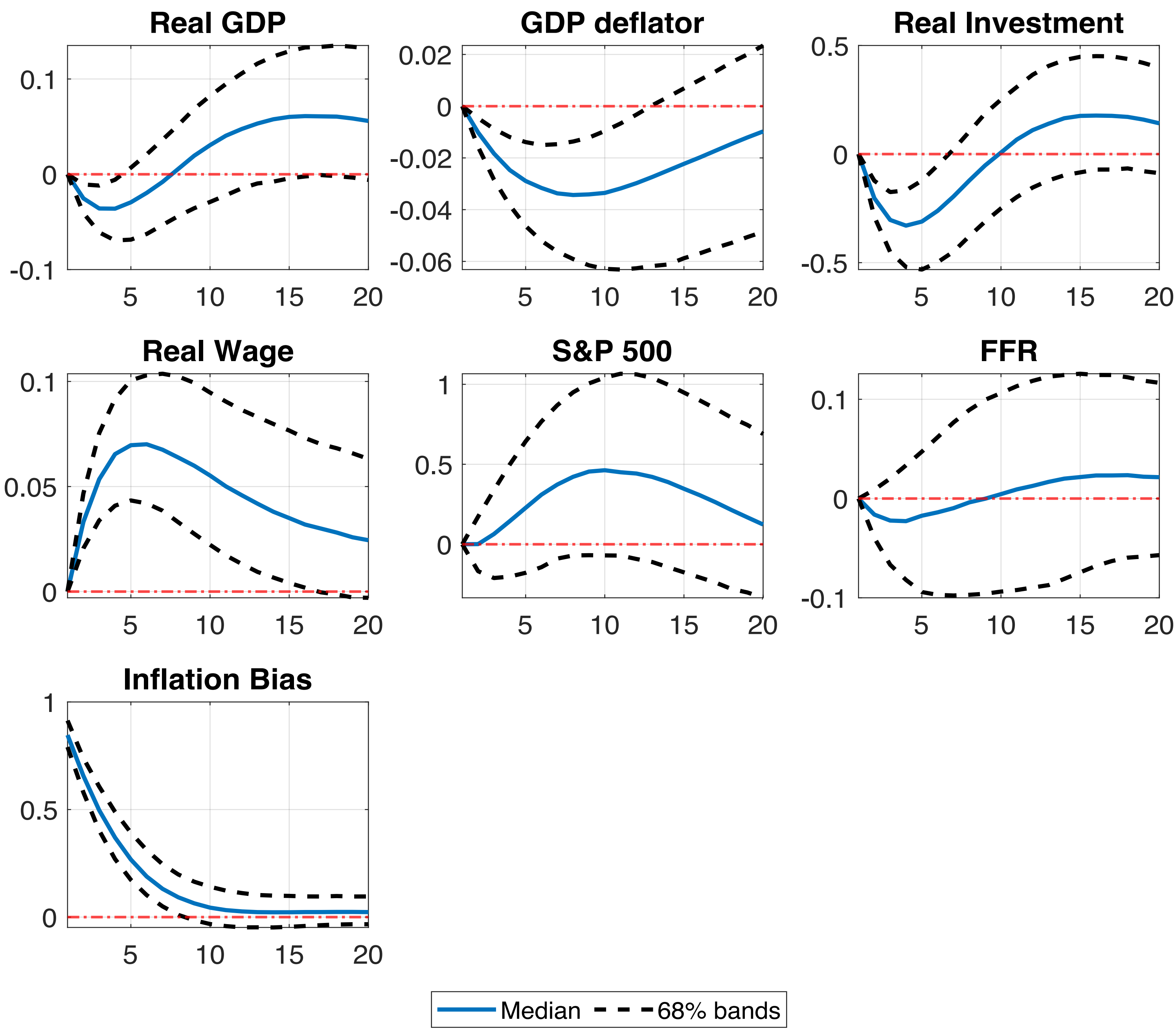
Impulse responses to a one standard deviation inflation bias index shock. Estimates from quarterly VAR with one lag. The bias index $\overline{bias}_t^{\pi}$ is constructed as the first principle component of inflation biases across all percentiles of SPF and BC. An increase in the bias means that forecasters systematically over-predict future inflation. Units are in percentage points. The sample is 1995:Q1-2018:Q2.

Figure 16: Responses to Inflation Bias Shock: SOC Only



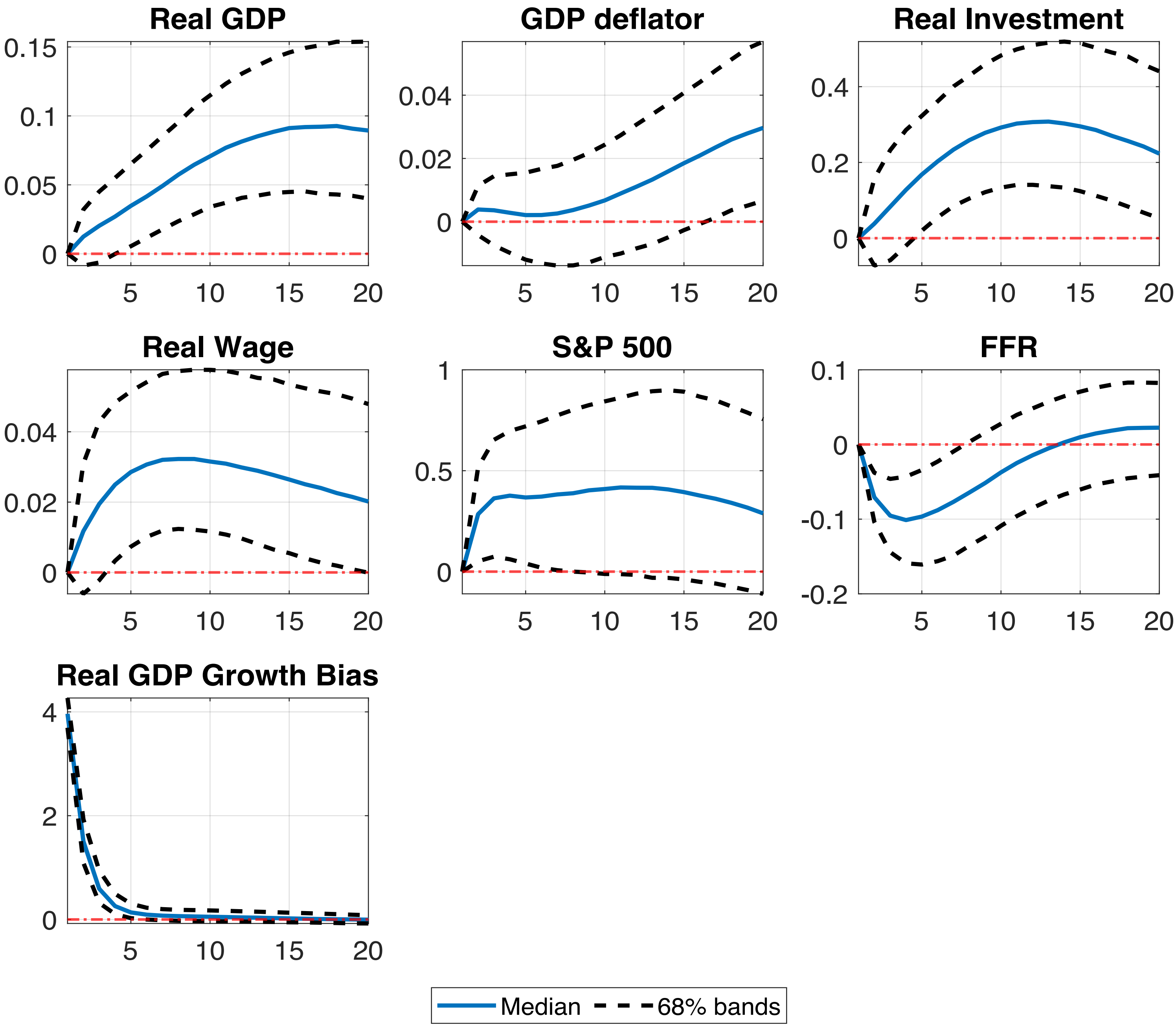
Impulse responses to a one standard deviation inflation bias index shock. Estimates from quarterly VAR with one lag. The bias index $\overline{bias}_t^{\pi}$ is constructed as the first principle component of inflation biases across all percentiles of SOC. An increase in the bias means that forecasters systematically over-predict future inflation. Units are in percentage points. The sample is 1995:Q1-2018:Q2.

Figure 17: Responses to Inflation Bias Shock: SPF and BC and SOC



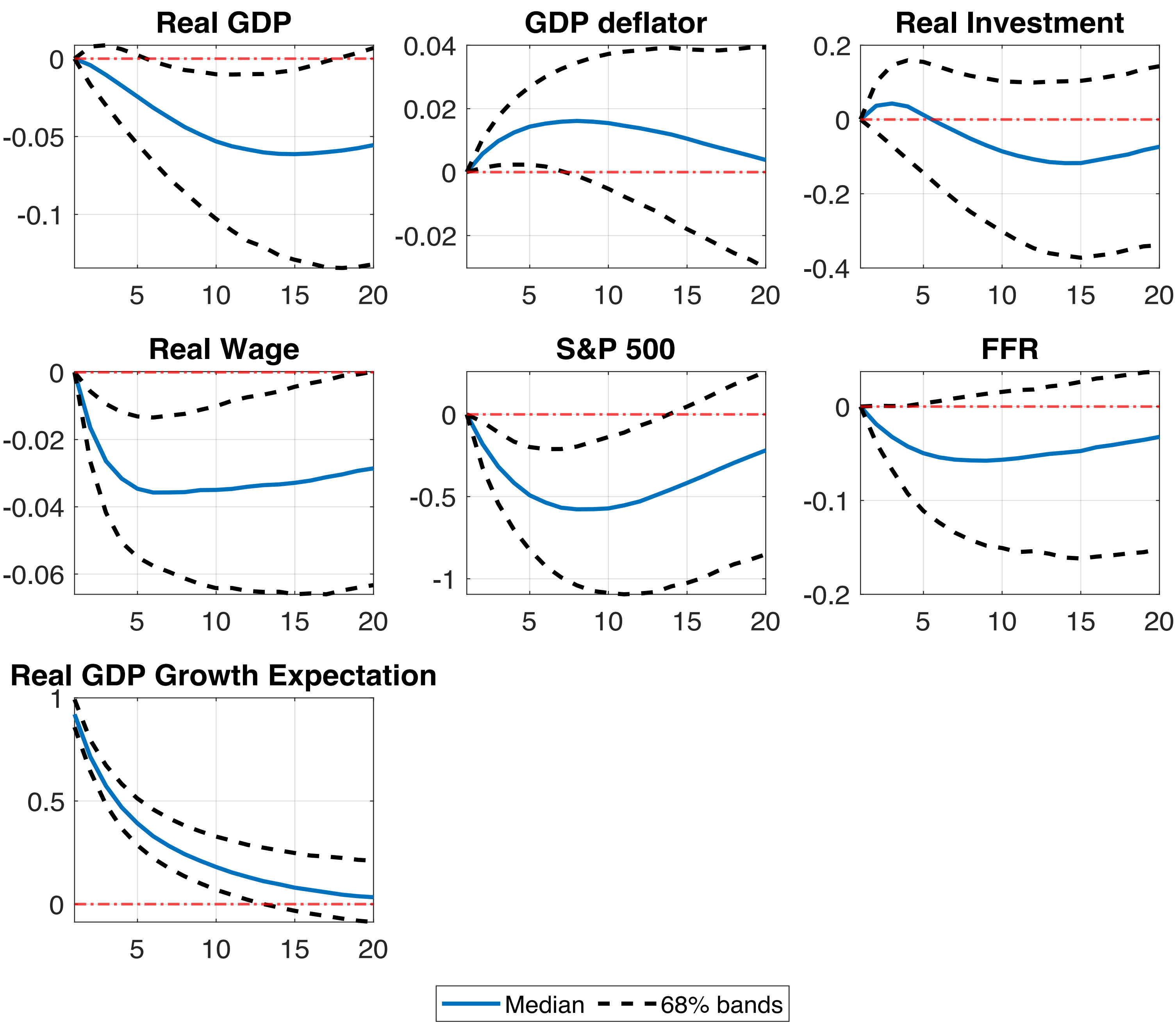
Impulse responses to a one standard deviation inflation bias index shock. Estimates from quarterly VAR with one lag. The bias index $\overline{bias}_t^{\pi}$ is constructed as the first principle component of inflation biases across all percentiles of SPF and BC and median of SOC. An increase in the bias means that forecasters systematically over-predict future inflation. Units are in percentage points. The sample is 1995:Q1-2018:Q2.

Figure 18: Responses to GDP Growth Bias Shock: SPF and BC and SOC



Impulse responses to a one standard deviation GDP growth bias index shock. Estimates from a quarterly VAR with one lag. The bias index $\overline{bias}_t^{\Delta y}$ is constructed as the first principle component of GDP growth biases across all surveys and percentiles. An increase in the bias means that forecasters overestimate future Real GDP growth. Units are percentage points. The sample is 1995:Q1-2018:Q2.

Figure 19: Responses to GDP Growth Expectation Shock



Impulse responses to a one standard deviation GDP growth expectation shock. Estimates from a quarterly VAR with one lag. The GDP growth expectation $\overline{\mathbb{F}}_t^{\Delta y}$ is constructed as the first principle component of GDP growth survey forecast across all surveys and percentiles. Units are percentage points. The sample is 1995:Q1-2018:Q2.

Online Appendix

Data

This appendix describes our data.

5.0.1 VAR Data

Real GDP: The real Gross Domestic Product is obtained from the US Bureau of Economic Analysis. It is in billions of chained 2012 dollars, quarterly frequency, seasonally adjusted, and at annual rate. We take the log of this variable. The source is from Bureau of Economic Analysis (BEA code: A191RX). The sample spans 1960:Q1 to 2019:Q3.

Real personal consumption expenditures: The real Personal Consumption Expenditures is obtained from the US Bureau of Economic Analysis. It is in billions of chained 2012 dollars, quarterly frequency, seasonally adjusted, and at annual rate. We take the log of this variable. The source is from Bureau of Economic Analysis (BEA code: DPCERX). The sample spans 1960:Q1 to 2019:Q3.

GDP price deflator: The Gross Domestic Product: implicit price deflator is obtained from the US Bureau of Economic Analysis. Index base is 2012=100, quarterly frequency, and seasonally adjusted. We take the log of this variable. The source is from Bureau of Economic Analysis (BEA code: A191RD). The sample spans 1960:Q1 to 2019:Q3.

Real investment: The real Gross Private Domestic Investment is obtained from the US Bureau of Economic Analysis. It is in billions of chained 2012 dollars, quarterly frequency, seasonally adjusted, and at annual rate. We take the log of this variable. The source is from Bureau of Economic Analysis (BEA code: A006RX). The sample spans 1960:Q1 to 2019:Q3.

Real wage: We obtain real wages by dividing the Average Hourly Earnings of Production and Nonsupervisory Employees: Manufacturing over the Personal Consumption Expenditures (implicit price deflator). Average Hourly Earnings of Production and Nonsupervisory Employees: Manufacturing is obtained from the US Bureau of Labor Statistics; it is in dollars per hour, quarterly frequency (average), and seasonally adjusted. BLS Account Code: CES3000000008. Personal Consumption Expenditures (implicit price deflator) is obtained from the US Bureau of Economic Analysis. Index base is 2012=100, quarterly frequency, and seasonally adjusted. We take the log of the ratio of these variables. The source is from Bureau of Economic Analysis (BEA code: DPCERD). The sample spans 1960:Q1 to 2019:Q3.

S&P 500 stock market index: The S&P 500 is obtained from the S&P Dow Jones Indices LLC. It is the quarterly average of the daily index value at market close. We take the log of this variable. The sample spans 1960:Q1 to 2019:Q3.

Federal funds rate (FFR): The Effective Federal Funds Rate is obtained from the Board of Governors of the Federal Reserve System. It is in percentage points, quarterly frequency (average), and not seasonally adjusted. The sample spans 1960:Q1 to 2019:Q3.

5.0.2 Survey Data

All details on survey data and survey forecast construction here, with links to data sources.

Survey of Professional Forecasters The SPF is conducted each quarter by sending out surveys to professional forecasters, defined as forecasters. The number of surveys sent varies over time, but recent waves sent around 50 surveys each quarter according to officials at the Federal Reserve Bank of Philadelphia. Only forecasters with sufficient academic training and experience as macroeconomic forecasters are eligible to participate. Over the course of our sample, the number of respondents ranges from a minimum of 9, to a maximum of 83, and the mean number of respondents is 37. The surveys are sent out at the end of the first month of each quarter, and they are collected in the second or third week of the middle month of each quarter. Each survey asks respondents to provide nowcasts and quarterly forecasts from one to four quarters ahead for a variety of variables. Specifically, we use the SPF micro data on individual forecasts of the price level, long-run inflation, and real GDP.[7] Below we provide the exact definitions of these variables as well as our method for constructing nowcasts and forecasts of quarterly and annual inflation and GDP growth for each respondent.[8]

The following variables are used on either the right- or left-hand-sides of forecasting models:

1. Quarterly and annual inflation (1968:Q4 - present): We use survey responses for the level of the GDP price index (PGDP), defined as

   "Forecasts for the quarterly and annual level of the chain-weighted GDP price index. Seasonally adjusted, index, base year varies. 1992-1995, GDP implicit deflator. Prior to 1992, GNP implicit deflator. Annual forecasts are for the annual average of the quarterly levels."

   Quarterly and annual inflation forecasts are constructed as follows. Let $\mathbb{F}_t^{(i)}\left[P_{t+h}\right]$ be forecaster i's prediction of PGDP h quarters ahead and $\mathbb{N}_t^{(i)}\left[P_t\right]$ be forecaster i's nowcast of PGDP for the current quarter. Annualized inflation forecasts for forecaster i are

   $$\mathbb{F}_t^{(i)}\left[\pi_{t+h,t}\right] = (400/h) \times \ln\left(\frac{\mathbb{F}_t^{(i)}\left[P_{t+h}\right]}{\mathbb{N}_t^{(i)}\left[P_t\right]}\right), \tag{A.11}$$

   where $h = 1$ for quarterly inflation and $h = 4$ for annual inflation. Similarly, we construct quarterly and annual nowcasts of inflation as

   $$\mathbb{N}_t^{(i)}\left[\pi_{t,t-h}\right] = (400/h) \times \ln\left(\frac{\mathbb{N}_t^{(i)}\left[P_t\right]}{P_{t-h}}\right),$$

   where $h = 1$ for quarterly inflation and $h = 4$ for annual inflation, and where P_{t-1} is the BEA's advance estimate of PGDP in the previous quarter observed by the respondent in time t, and P_{t-4} is the BEA's most accurate estimate of PGDP four quarters back. After computing inflation for each survey respondent, we calculate the 5th through the

[7] Individual forecasts for all variables can be downloaded at https://www.philadelphiafed.org/research-and-data/real-time-center/survey-of-professional-forecasters/historical-data/individual-forecasts.

[8] The SPF documentation file can be found at https://www.philadelphiafed.org/-/media/research-and-data/real-time-center/survey-of-professional-forecasters/spf-documentation.pdf?la=en.

95th percentiles as well as the average, variance, and skewness of inflation forecasts across respondents.

2. Long-run inflation (1991:Q4 - present): We use survey responses for 10-year-ahead CPI inflation (CPI10), which is defined as

   "Forecasts for the annual average rate of headline CPI inflation over the next 10 years. Seasonally adjusted, annualized percentage points. The "next 10 years" includes the year in which we conducted the survey and the following nine years. Conceptually, the calculation of inflation is one that runs from the fourth quarter of the year before the survey to the fourth quarter of the year that is ten years beyond the survey year, representing a total of 40 quarters or 10 years. The fourth-quarter level is the quarterly average of the underlying monthly levels."

   Only the median response is provided for CPI10, and it is already reported as an inflation rate, so we do not make any adjustments and cannot compute other moments or percentiles.

3. Real GDP growth (1968:Q4 - present): We use the level of real GDP (RGDP), which is defined as

   *"Forecasts for the quarterly and annual level of chain-weighted real GDP. Seasonally adjusted, annual rate, base year varies. 1992-1995, fixed-weighted real GDP. Prior to 1992, fixed-weighted real GNP. Annual forecasts are for the annual average of the quarterly levels. Prior to 1981:Q3, RGDP is computed by using the formula NGDP / PGDP * 100."*

   Quarterly and annual growth rates are constructed the same way as for inflation, except RGDP replaces PGDP.

In order to generate OOS forecasts that could have been made in real time, it is necessary to take a stand on the information set of the forecasters when each forecast was made. We assume that forecasters could have used all data released before the survey deadlines. Table A.1 lists the survey deadlines that are available, beginning with the 1990:Q3 survey. Before 1990:Q3, we make the conservative assumption that respondents only had data released by the first day of the second month of each quarter.

Table A.1: SPF Survey Deadlines[9]

Survey	Deadline Date	Survey	Deadline Date	Survey	Deadline Date
1990:Q1	NA	1991:Q1	2/16/91	1992:Q1	2/22/92
Q2	NA	Q2	5/18/91	Q2	5/15/92
Q3	8/23/90	Q3	8/18/91	Q3	8/21/92
Q4	11/22/90	Q4	11/16/91	Q4	11/20/92
1993:Q1	2/19/93	1994:Q1	2/21/94	1995:Q1	2/21/95
Q2	5/20/93	Q2	5/18/94	Q2	5/22/95
Q3	8/19/93	Q3	8/18/94	Q3	8/22/95
Q4	11/23/93	Q4	11/18/94	Q4	11/20/95
1996:Q1	3/2/96	1997:Q1	2/19/97	1998:Q1	2/18/98

[9]SPF survey deadlines are posted online at https://www.philadelphiafed.org/-/media/research-and-data/real-time-center/survey-of-professional-forecasters/spf-release-dates.txt?la=en.

Table A.1 (Cont'd)

Survey	Deadline Date	Survey	Deadline Date	Survey	Deadline Date
Q2	5/18/96	Q2	5/17/97	Q2	5/16/98
Q3	8/21/96	Q3	8/16/97	Q3	8/15/98
Q4	11/18/96	Q4	11/19/97	Q4	11/14/98
1999:Q1	2/16/99	2000:Q1	2/12/00	2001:Q1	2/14/01
Q2	5/15/99	Q2	5/13/00	Q2	5/12/01
Q3	8/14/99	Q3	8/12/00	Q3	8/15/01
Q4	11/13/99	Q4	11/11/00	Q4	11/14/01
2002:Q1	2/12/02	2003:Q1	2/14/03	2004:Q1	2/14/04
Q2	5/13/02	Q2	5/12/03	Q2	5/14/04
Q3	8/14/02	Q3	8/16/03	Q3	8/13/04
Q4	11/13/02	Q4	11/14/03	Q4	11/13/04
2005:Q1	2/9/05	2006:Q1	2/8/06	2007:Q1	2/8/07
Q2	5/12/05	Q2	5/10/06	Q2	5/9/07
Q3	8/11/05	Q3	8/9/06	Q3	8/8/07
Q4	11/8/05	Q4	11/8/06	Q4	11/7/07
2008:Q1	2/7/08	2009:Q1	2/10/09	2010:Q1	2/9/10
Q2	5/8/08	Q2	5/12/09	Q2	5/11/10
Q3	8/7/08	Q3	8/11/09	Q3	8/10/10
Q4	11/10/08	Q4	11/10/09	Q4	11/9/10
2011:Q1	2/8/11	2012:Q1	2/7/12	2013:Q1	2/11/13
Q2	5/10/11	Q2	5/8/12	Q2	5/7/13
Q3	8/8/11	Q3	8/7/12	Q3	8/12/13
Q4	11/8/11	Q4	11/6/12	Q4	11/18/13
2014:Q1	2/10/14	2015:Q1	2/10/15	2016:Q1	2/9/16
Q2	5/11/14	Q2	5/12/15	Q2	5/10/16
Q3	8/11/14	Q3	8/11/15	Q3	8/9/16
Q4	11/10/14	Q4	11/10/15	Q4	11/8/16
2017:Q1	2/7/17	2018:Q1	2/6/18		
Q2	5/9/17	Q2	5/8/18		
Q3	8/8/17	Q3	8/7/18		
Q4	11/7/17	Q4	11/6/18		

Michigan Survey of Consumers (SOC) We construct MS forecasts of annual inflation and GDP growth of respondents answering at time t. Each month, the SOC contains approximately 50 core questions, and a minimum of 500 interviews are conducted by telephone over the course of the entire month, each month. We use two questions from the monthly survey for which the time series begins in January 1978, and convert to quarterly observations as explained below.

1. Annual CPI inflation: We use survey responses to question A12b, which asks (emphasis in original):

   By about what percent do you expect prices to go (up/down) on the average, during the next 12 months?

   Respondents provide a numerical value to the interviewer and the SOC provides the mean, median, and 25th and 75th percentiles. Since this is already reported as an inflation rate, we do not make any adjustments.

2. Annual real GDP growth: We use survey responses to question A7, which asks (emphasis in original):

And how about a year from now, do you expect that in the country as a whole business conditions will be <u>better</u>*, or* <u>worse</u> *than they are at present, or just about the same?*

Respondents select one of three options: "better a year from now," "about the same," or "worse a year from now." There is a long history of using survey data as a proxy for spending and output (see, for example, Ludvigson - "Consumer Confidence and Consumer Spending" - Journal of Economic Perspectives - 2004). Using a companion question in the SOC that asks about contemporaneous business conditions, Curtin (2019) and the SOC survey documentation suggest constructing a "balance score" to generate a contemporaneous measure of real GDP growth. The *balance score* equals the percentage of respondents who expected that the economy to improve minus the percentage that expected it to worsen + 100. Applying this methodology to question A7.

The balance score is obtained monthly and we use the observation for the middle month of each quarter as our quarterly observation. We convert the score to a quantitative survey-based measure of real GDP growth using a simple linear regression. Specifically, at time s, we assume that GDP growth, $y_{j,s+4}$, is related to the contemporaneous Michigan Survey balance sore, M_s, by:

$$y_{j,s+4} = \beta_0 + \beta_1 M_s + \epsilon_s.$$

This equation is estimated using OLS and the real-time vintage data, and then the forecast is constructed as $\mathbb{F}_{j,t}[y_{j,t+4}] = \hat{\beta}_0 + \hat{\beta}_1 M_t$

Specifically, we first estimate the coefficients of this regression over the sample 1978:Q1-1994:Q1. Using the estimated coefficients and the balance score from 1995:Q1 gives us the point forecast of inflation for 1995:Q1-1996:Q1. We then re-estimate this equation, recursively, adding one observation to the end of the sample at a time, and storing the fitted values. This results in a time series of forecasts $\mathbb{F}_{j,t}[y_{j,t+4}]$.

As with the SPF, we take a stand on the information set of consumers when each forecast was made, and we assume that consumers could have used all data released before they completed the survey. For the SOC interviews are conducted monthly over the course of an entire month. We set the interview response deadline for each survey as the first day of the survey month. For example, we set the deadline to February 1st, 2019, for the February 2019 Survey of Consumers, while in reality, the interview period was from February 2 to February 29, 2019. In other months, the true interview start period may be near the end of the previous month, such as in February 2019, when it was January 31st, 2019. To align the SOC more closely with the SPF deadline for survey completion (end of the second or third week of the middle month of the quarter), we use the middle month of each quarter as our quarterly observation for the SOC.

Bluechip Data We obtain Blue Chip expectation data from Blue Chip Financial Forecasts. The surveys are conducted each month by sending out surveys to forecasters in around 50 financial firms such as Bank of America, Goldman Sachs & Co., Swiss Re, Loomis, Sayles & Company, and J.P. Morgan Chase. The participants are surveyed around the 25th of each month and the results published a few days later on the 1st of the following month. The forecasters are asked to forecast the average of the level of U.S. interest rates over a particular calendar quarter, e.g. the federal funds rate and the set of H.15 Constant Maturity Treasuries (CMT) of the following maturities: 3-month, 6-month, 1-year, 2-year, 5-year and 10-year, and the quarter over quarter percentage changes in Real GDP, the GDP Price Index and the Consumer Price Index, beginning with the current quarter and extending 4 to 5 quarters into the future.

In this study, we look at a subset of the forecasted variables. Specifically, we use the Blue Chip micro data on individual forecasts of the quarter-over-quarter (Q/Q) percentage change in the Real GDP, the GDP Price Index and the CPI, and convert to quarterly observations as explained below.

1. Quarterly and annual PGDP inflation (1986:Q1 - 2018:Q3): We use survey responses for the quarter-over-quarter percentage change in the GDP price index, defined as:

   "Forecasts for the quarter-over-quarter percentage change in the GDP Chained Price Index. Seasonally adjusted annual rate (SAAR). 1992 Jan. to 1996 June, Q/Q % change (SAAR) in GDP implicit deflator. 1986 Jan. to 1991 Dec., Q/Q % change (SAAR) in GNP implicit deflator."

   Quarterly and annual inflation forecasts are constructed as follows. Let $\mathbb{F}_t^{(i)}\left[gP_{t+h}^{(Q/Q)}\right]$ be forecaster i's prediction of Q/Q % change in PGDP h quarters ahead. Annualized inflation forecasts for forecaster i in the next quarter are:

   $$\mathbb{F}_t^{(i)}\left[\pi_{t+1,t}\right] = 400 \times \ln\left(1 + \frac{\mathbb{F}_t^{(i)}\left[gP_{t+1}^{(Q/Q)}\right]}{100}\right)^{\frac{1}{4}}$$

   Annual Inflation forecasts are:

   $$\mathbb{F}_t^{(i)}\left[\pi_{t+4,t}\right] = 100 \times \ln\left(\prod_{h=1}^{4}\left(1 + \frac{\mathbb{F}_t^{(i)}\left[gP_{t+h}^{(Q/Q)}\right]}{100}\right)\right)^{\frac{1}{4}}$$

   Quarterly nowcasts of inflation are constructed as:

   $$\mathbb{N}_t^{(i)}\left[\pi_{t,t-1}\right] = 400 \times \ln\left(1 + \frac{\mathbb{N}_t^{(i)}\left[gP_{t}^{(Q/Q)}\right]}{100}\right)^{\frac{1}{4}}$$

   where $\mathbb{N}_t^{(i)}\left[gP_t^{(Q/Q)}\right]$ is forecaster i's nowcast of Q/Q % change in PGDP for the current quarter. Annual nowcasts of inflation for forecaster i are:

   $$\mathbb{N}_t^{(i)}\left[\pi_{t,t-4}\right] = 100 \times \ln\left(\frac{\mathbb{N}_t^{(i)}\left[P_t\right]}{P_{t-4}}\right),$$

   where P_{t-4} is the BEA's most accurate estimate of PGDP four quarters back and $\mathbb{N}_t^{(i)}\left[P_t\right]$ is forecaster i's nowcast of PGDP for the current quarter which is constructed as: $\mathbb{N}_t^{(i)}\left[P_t\right] = \exp\left(\mathbb{N}_t^{(i)}\left[\pi_{t,t-1}\right]/400 + \ln P_{t-1}\right)$. Similarly, we also calculate the 5th through the 95th percentiles as well as the average, variance, and skewness of inflation forecasts across respondents.

2. Real GDP growth (1984:Q3 - 2018:Q3): We use quarter-over-quarter percentage change in the Real GDP, which is defined as

   "Forecasts for the quarter-over-quarter percentage change in the level of chain-weighted real GDP. Seasonally adjusted, annual rate. Prior to 1992, Q/Q % change (SAAR) in real GNP."

   Quarterly and annual growth rates are constructed the same way as for inflation, except RGDP replaces PGDP.

3. CPI inflation (1984:Q3 - 2018:Q3): We use quarter-over-quarter percentage change in the consumer price index, which is defined as

   "Forecasts for the quarter-over-quarter percentage change in the CPI (consumer prices for all urban consumers). Seasonally adjusted, annual rate."

   Quarterly and annual CPI inflation are constructed the same way as for PGDP inflation, except CPI replaces PGDP.

The surveys are conducted right before the publication of the newsletter. Each issue is always dated the 1st of the month and the actual survey conducted over a two-day period almost always between 24th and 28th of the month. The major exception is the January issue when the survey is conducted a few days earlier to avoid conflict with the Christmas holiday. Therefore, we assume that the end of the last month (equivalently beginning of current month) is when the forecast is made. For example, for the report in 2008 Feb, we assume that the forecast is made on Feb 1, 2008. To convert monthly forecasts to quarterly forecasts, we use the forecasts in the middle month of each quarter as the quarterly forecasts. This is to align the Blue Chip more closely with the SPF deadline for survey completion, similar to what we do for the SOC.

Real-Time Macro Data

At each forecast date in the sample, we construct a dataset of macro variables that could have been observed on or before the day of the survey deadline. We use the Philadelphia Fed's Real-Time Data Set to obtain vintages of macro variables.[10] These vintages capture changes to historical data due to periodic revisions made by government statistical agencies. The vintages for a particular series can be available at the monthly and/or quarterly frequencies, and the series have monthly and/or quarterly observations. In cases where a variable has both frequencies available for its vintages and/or its observations, we choose one format of the variable. For instance, nominal personal consumption expenditures on goods is quarterly data with both monthly and quarterly vintages available; in this case, we use the version with monthly vintages.

Real Time Regressands Following CG, all regressions are run and forecast errors computed using forecasts of real-time inflation and GDP data available four quarters after the period being

[10]The real-time data sets are available at https://www.philadelphiafed.org/research-and-data/real-time-center/real-time-data/data-files.

forecast. Following Faust and Wright (2013), we use continuous time compounding of inflation and GDP growth. For example, four quarter inflation is computed as

$$\pi_{t+4,t} = (100) \times \ln\left(\frac{P_{t+h}}{P_t}\right),$$

where P_t is the time t price level.

Real Time Regressors For the regressors we need to combine all of the data observed at the time of a forecast date, and know the specific day that the data in each vintage are released. It is not sufficient to align vintage dates with forecast dates because the time t vintage might include data released after the time t forecast was made. The series-specific documentation on the Philadelphia Fed's website provides details on the timing of the vintages for each series. For some series, exact release dates are known, and thus the vintages reflect the data available at the time of the data release. When this is the case, we download the release dates from the relevant statistical agency and compare each vintage release date to the corresponding survey deadline to determine whether a particular vintage can be included in a survey respondent's information set.

For other variables, we only know that vintages contain data available in the middle of a month or quarter, but not the exact day. A subset of these variables come from the BEA National Income and Product Accounts, which are released at the end of each month. Since NIPA series are released at the end of each month, and vintages reflect data available in the middle of each month, a survey respondent making a forecast in the middle of a month includes the current month's vintage of NIPA data in her information set. However, there is another subset of variables with unknown release dates, for which we must make the conservative assumption that a forecaster at time t observes at most the time $t-1$ vintage of data. An Excel Workbook containing the known release dates and timing assumptions is available on the authors' websites.

In addition to the macro variables with different vintages that we obtain from the Philadelphia Fed, we include energy prices from the U.S. Bureau of Labor Statistics (BLS). Energy prices do not get revised, so they do not have multiple vintages. Instead there is just one historical version of the data.

After combining all of the series that are known by the forecasters at each date, we convert monthly data to quarterly by using either the beginning-of-quarter or end-of-quarter values. The decision to use beginning-of-quarter or end-of-quarter depends on the survey deadline of a particular forecast date. If the survey deadline is known to be in the middle of the second month of quarter t, then it is conceivable that the forecasters would have information about the first month of quarter t. Therefore, we use beginning-of-quarter values. Alternatively, if the survey deadline is unknown we allow only information up to quarter $t-1$ to enter the model. Thus, we use end-of-quarter values in these cases.

Table A.2 lists the Philadelphia Fed variables as well as the energy prices data from the BLS. Included in the table is the first available vintages for each variable that has multiple vintages. We do not include the last vintage because most variables have vintages through the present.[11] Table A.2 also lists the transformation applied to each variable to make them stationary before

[11] For variables BASEBASAQVMD, NBRBASAQVMD, NBRECBASAQVMD, and TRBASAQVMD, the last available vintage is 2013:Q2.

generating factors. Let X_{it} denote variable i at time t after the transformation, and let X_{it}^A be the untransformed series. Let $\Delta = (1 - L)$ with $LX_{it} = X_{it-1}$. There are seven possible transformations with the following codes:

1 Code lv: $X_{it} = X_{it}^A$
2 Code Δlv: $X_{it} = X_{it}^A - X_{it-1}^A$
3 Code $\Delta^2 lv$: $X_{it} = \Delta^2 X_{it}^A$
4 Code ln: $X_{it} = \ln(X_{it}^A)$
5 Code Δln: $X_{it} = \ln(X_{it}^A) - \ln(X_{it-1}^A)$
6 Code $\Delta^2 ln$: $X_{it} = \Delta^2 \ln(X_{it}^A)$
7 Code $\Delta lv/lv$: $X_{it} = (X_{it}^A - X_{it-1}^A)/X_{it-1}^A$

Table A.2: List of Macro Dataset Variables

No.	Short Name	Source	Tran	Description	First Vintage
			Group 1: Output and Income		
1	IPMMVMD	Philly Fed	Δln	Ind. production index - Manufacturing	1962:M11
2	IPTMVMD	Philly Fed	Δln	Ind. production index - Total	1962:M11
3	CUMMVMD	Philly Fed	lv	Capacity utilization - Manufacturing	1979:M8
4	CUTMVMD	Philly Fed	lv	Capacity utilization - Total	1983:M7
5	NCPROFATMVQD	Philly Fed	Δln	Nom. corp. profits after tax without IVA/CCAdj	1965:Q4
6	NCPROFATWMVQD	Philly Fed	Δln	Nom. corp. profits after tax with IVA/CCAdj	1981:Q1
7	OPHMVQD	Philly Fed	Δln	Output per hour - Business sector	1998:Q4
8	NDPIQVQD	Philly Fed	Δln	Nom. disposable personal income	1965:Q4
9	NOUTPUTQVQD	Philly Fed	Δln	Nom. GNP/GDP	1965:Q4
10	NPIQVQD	Philly Fed	Δln	Nom. personal income	1965:Q4
11	NPSAVQVQD	Philly Fed	Δlv	Nom. personal saving	1965:Q4
12	OLIQVQD	Philly Fed	Δln	Other labor income	1965:Q4
13	PINTIQVQD	Philly Fed	Δln	Personal interest income	1965:Q4
14	PINTPAIDQVQD	Philly Fed	Δln	Interest paid by consumers	1965:Q4
15	PROPIQVQD	Philly Fed	Δln	Proprietors' income	1965:Q4
16	PTAXQVQD	Philly Fed	Δln	Personal tax and nontax payments	1965:Q4
17	RATESAVQVQD	Philly Fed	Δlv	Personal saving rate	1965:Q4
18	RENTIQVQD	Philly Fed	Δlv	Rental income of persons	1965:Q4
19	ROUTPUTQVQD	Philly Fed	Δln	Real GNP/GDP	1965:Q4
20	SSCONTRIBQVQD	Philly Fed	Δln	Personal contributions for social insurance	1965:Q4
21	TRANPFQVQD	Philly Fed	Δln	Personal transfer payments to foreigners	1965:Q4
22	TRANRQVQD	Philly Fed	Δln	Transfer payments	1965:Q4
23	CUUR0000SA0E	BLS	$\Delta^2 ln$	Energy in U.S. city avg., all urban consumers, not seasonally adj	
			Group 2: Employment		
24	EMPLOYMVMD	Philly Fed	Δln	Nonfarm payroll	1946:M12
25	HMVMD	Philly Fed	lv	Aggregate weekly hours - Total	1971:M9
26	HGMVMD	Philly Fed	lv	Agg. weekly hours - Goods-producing	1971:M9
27	HSMVMD	Philly Fed	lv	Agg. weekly hours - Service-producing	1971:M9
28	LFCMVMD	Philly Fed	Δln	Civilian labor force	1998:M11
29	LFPARTMVMD	Philly Fed	lv	Civilian participation rate	1998:M11
30	POPMVMD	Philly Fed	Δln	Civilian noninstitutional population	1998:M11
31	ULCMVQD	Philly Fed	Δln	Unit labor costs - Business sector	1998:Q4
32	RUCQVMD	Philly Fed	Δlv	Unemployment rate	1965:Q4
33	WSDQVQD	Philly Fed	Δln	Wage and salary disbursements	1965:Q4
			Group 3: Orders and Investment		
34	HSTARTSMVMD	Philly Fed	Δln	Housing starts	1968:M2
35	RINVBFMVQD	Philly Fed	Δln	Real gross private domestic inv. - Nonresidential	1965:Q4
36	RINVCHIMVQD	Philly Fed	Δlv	Real gross private domestic inv. - Change in private inventories	1965:Q4
37	RINVRESIDMVQD	Philly Fed	Δln	Real gross private domestic inv. - Residential	1965:Q4
			Group 4: Consumption		
38	NCONGMMVMD	Philly Fed	Δln	Nom. personal cons. exp. - Goods	2009:M8
39	NCONHHMMVMD	Philly Fed	Δln	Nom. hh. cons. exp.	2009:M8
40	NCONSHHMMVMD	Philly Fed	Δln	Nom. hh. cons. exp. - Services	2009:M8

Table A.2 (Cont'd)

No.	Short Name	Source	Tran	Description	First Vintage
41	NCONSNPMMVMD	Philly Fed	Δln	Nom. final cons. exp. of NPISH	2009:M8
42	RCONDMMVMD	Philly Fed	Δln	Real personal cons. exp. - Durables	1998:M11
43	RCONGMMVMD	Philly Fed	Δln	Real personal cons. exp. - Goods	2009:M8
44	RCONHHMMVMD	Philly Fed	Δln	Real hh. cons. exp.	2009:M8
45	RCONMMVMD	Philly Fed	Δln	Real personal cons. exp. - Total	1998:M11
46	RCONNDMVMD	Philly Fed	Δln	Real personal cons. exp. - Nondurables	1998:M11
47	RCONSHHMMVMD	Philly Fed	Δln	Real hh. cons. exp. - Services	2009:M8
48	RCONSMMVMD	Philly Fed	Δln	Real personal cons. exp. - Services	1998:M11
49	RCONSNPMMVMD	Philly Fed	Δln	Real final cons. exp. of NPISH	2009:M8
50	NCONGMVQD	Philly Fed	Δln	Nom. personal cons. exp. - Goods	2009:Q3
51	NCONHHMVQD	Philly Fed	Δln	Nom. hh. cons. exp.	0209:Q3
52	NCONSHHMVQD	Philly Fed	Δln	Nom. hh. cons. exp. - Services	2009:Q3
53	NCONSNPMVQD	Philly Fed	Δln	Nom. final cons. exp. of NPISH	2009:Q3
54	RCONDMVQD	Philly Fed	Δln	Real personal cons. exp. - Durable goods	1965:Q4
55	RCONGMVQD	Philly Fed	Δln	Real personal cons. exp. - Goods	2009:Q3
56	RCONHHMVQD	Philly Fed	Δln	Real hh. cons. exp.	2009:Q3
57	RCONMVQD	Philly Fed	Δln	Real personal cons. exp. - Total	1965:Q4
58	RCONNDMVQD	Philly Fed	Δln	Real pesonal cons. exp. - Nondurable goods	1965:Q4
59	RCONSHHMVQD	Philly Fed	Δln	Real hh. cons. exp. - Services	2009:Q3
60	RCONSMVQD	Philly Fed	Δln	Real personal cons. exp. - Services	1965:Q4
61	RCONSNPMVQD	Philly Fed	Δln	Real final cons. exp. of NPISH	2009:Q3
62	NCONQVQD	Philly Fed	Δln	Nom. personal cons. exp.	1965:Q4
			Group 5: Prices		
63	PCONGMMVMD	Philly Fed	$\Delta^2 ln$	Price index for personal cons. exp. - Goods	2009:M8
64	PCONHHMMVMD	Philly Fed	$\Delta^2 ln$	Price index for hh. cons. exp.	2009:M8
65	PCONSHHMMVMD	Philly Fed	$\Delta^2 ln$	Price index for hh. cons. exp. - Services	2009:M8
66	PCONSNPMMVMD	Philly Fed	$\Delta^2 ln$	Price index for final cons. exp. of NPISH	2009:M8
67	PCPIMVMD	Philly Fed	$\Delta^2 ln$	Consumer price index	1998:M11
68	PCPIXMVMD	Philly Fed	$\Delta^2 ln$	Core consumer price index	1998:M11
69	PPPIMVMD	Philly Fed	$\Delta^2 ln$	Producer price index	1998:M11
70	PPPIXMVMD	Philly Fed	$\Delta^2 ln$	Core producer price index	1998:M11
71	PCONGMVQD	Philly Fed	$\Delta^2 ln$	Price index for personal. cons. exp. - Goods	2009:Q3
72	PCONHHMVQD	Philly Fed	$\Delta^2 ln$	Price index for hh. cons. exp.	2009:Q3
73	PCONSHHMVQD	Philly Fed	$\Delta^2 ln$	Price index for hh. cons. exp. - Services	2009:Q3
74	PCONSNPMVQD	Philly Fed	$\Delta^2 ln$	Price index for final cons. exp. of NPISH	2009:Q3
75	PCONXMVQD	Philly Fed	$\Delta^2 ln$	Core price index for personal cons. exp.	1996:Q1
76	CPIQVMD	Philly Fed	$\Delta^2 ln$	Consumer price index	1994:Q3
77	PQVQD	Philly Fed	$\Delta^2 ln$	Price index for GNP/GDP	1965:Q4
78	PCONQVQD	Philly Fed	$\Delta^2 ln$	Price index for personal cons. exp.	1965:Q4
79	PIMPQVQD	Philly Fed	$\Delta^2 ln$	Price index for imports of goods and services	1965:Q4
			Group 6: Trade and Government		
80	REXMVQD	Philly Fed	Δln	Real exports of goods and services	1965:Q4
81	RGMVQD	Philly Fed	Δln	Real government cons. and gross inv. - Total	1965:Q4
82	RGFMVQD	Philly Fed	Δln	Real government cons. and gross inv. - Federal	1965:Q4
83	RGSLMVQD	Philly Fed	Δln	Real government cons. and gross. inv. - State and local	1965:Q4
84	RIMPMVQD	Philly Fed	Δln	Real imports of goods and services	1965:Q4
85	RNXMVQD	Philly Fed	Δlv	Real net exports of goods and services	1965:Q4
			Group 7: Money and Credit		
86	BASEBASAQVMD	Philly Fed	$\Delta^2 ln$	Monetary base	1980:Q2
87	M1QVMD	Philly Fed	$\Delta^2 ln$	M1 money stock	1965:Q4
88	M2QVMD	Philly Fed	$\Delta^2 ln$	M2 money stock	1971:Q2
89	NBRBASAQVMD	Philly Fed	$\Delta lv/lv$	Nonborrowed reserves	1967:Q3
80	NBRECBASAQVMD	Philly Fed	$\Delta lv/lv$	Nonborrowed reserves plus extended credit	1984:Q2
91	TRBASAQVMD	Philly Fed	$\Delta^2 ln$	Total reserves	1967:Q3
92	DIVQVQD	Philly Fed	Δln	Dividends	1965:Q4

Monthly Financial Factor Data

The 147 financial series in this data set are versions of the financial dataset used in Jurado, Ludvigson, and Ng (2015) and Ludvigson, Ma, and Ng (2019). It consists of a number of indicators measuring the behavior of a broad cross-section of asset returns, as well as some aggregate financial indicators not included in the macro dataset. These data include valuation ratios such as the dividend-price ratio and earnings-price ratio, growth rates of aggregate dividends and prices, default and term spreads, yields on corporate bonds of different ratings grades, yields on Treasuries and yield spreads, and a broad cross-section of industry equity returns. Following Fama and French (1992), returns on 100 portfolios of equities sorted into 10 size and 10 book-market categories. The dataset X^f also includes a group of variables we call "risk-factors," since they have been used in cross-sectional or time-series studies to uncover variation in the market risk-premium. These risk-factors include the three Fama and French (1993) risk factors, namely the excess return on the market MKT_t, the "small-minus-big" (SMB_t) and "high-minus-low" (HML_t) portfolio returns, the momentum factor UMD_t, and the small stock value spread $R15 - R11$.

The raw data used to form factors are always transformed to achieve stationarity. In addition, when forming forecasting factors from the large macro and financial datasets, the raw data (which are in different units) are standardized before performing PCA. When forming common uncertainty from estimates of individual uncertainty, the raw data (which are in this case in the same units) are demeaned, but we do not divide by the observation's standard deviation before performing PCA.

Throughout, the factors are estimated by the method of static principal components (PCA). Specifically, the $T \times r_F$ matrix $\hat{F}_t$ is $\sqrt{T}$ times the r_F eigenvectors corresponding to the r_F largest eigenvalues of the $T \times T$ matrix $xx'/(TN)$ in decreasing order. In large samples (when $\sqrt{T}/N \rightarrow \infty$), Bai and Ng (2006) show that the estimates $\hat{F}_t$ can be treated as though they were observed in the subsequent forecasting regression.

All returns and spreads are expressed in logs (i.e. the log of the gross return or spread), are displayed in percent (i.e. multiplied by 100), and are annualized by multiplying by 12, i.e., if x is the original return or spread, we transform to $1200 \ln(1 + x/100)$. Federal Reserve data are annualized by default and are therefore not "re-annualized." Note: this annualization means that the annualized standard deviation (volatility) is equal to the data standard deviation divided by $\sqrt{12}$. The data series used in this dataset are listed below by data source. Additional details on data transformations are given below the table.

Let X_{it} denote variable i observed at time t after e.g., logarithm and differencing transformation, and let X_{it}^A be the actual (untransformed) series. Let $\Delta = (1 - L)$ with $LX_{it} = X_{it-1}$. There are six possible transformations with the following codes:

1 Code lv: $X_{it} = X_{it}^A$.

2 Code Δlv: $X_{it} = X_{it}^A - X_{it-1}^A$.

3 Code $\Delta^2 lv$: $X_{it} = \Delta^2 X_{it}^A$.

4 Code ln: $X_{it} = ln(X_{it}^A)$.

5 Code Δln: $X_{it} = ln(X_{it}^A) - ln(X_{it-1}^A)$.

6 Code $\Delta^2 ln$: $X_{it} = \Delta^2 ln X_{it}^A$.

7 Code $\Delta lv/lv$: $\left(X_{it}^A - X_{it-1}^A\right)/X_{it-1}^A$

Table A.3: List of Financial Dataset Variables

No.	Short Name	Source	Tran	Description
		Group 1: Prices, Yield, Dividends		
1	D_log(DIV)	CRSP	Δln	$\Delta \log D_t^*$ see additional details below
2	D_log(P)	CRSP	Δln	$\Delta \log P_t$ see additional details below
3	D_DIVreinvest	CRSP	Δln	$\Delta \log D_t^{re,*}$ see additional details below
4	D_Preinvest	CRSP	Δln	$\Delta \log P_t^{re,*}$ see additional details below
5	d-p	CRSP	ln	$\log(D_t^*) - \log P_t$ see additional details below
		Group 2: Equity Risk Factors		
6	R15-R11	Kenneth French	lv	(Small, High) minus (Small, Low) sorted on (size, book-to-market)
7	Mkt-RF	Kenneth French	lv	Market excess return
8	SMB	Kenneth French	lv	Small Minus Big, sorted on size
9	HML	Kenneth French	lv	High Minus Low, sorted on book-to-market
10	UMD	Kenneth French	lv	Up Minus Down, sorted on momentum
		Group 3: Industries		
11	Agric	Kenneth French	lv	Agric industry portfolio
12	Food	Kenneth French	lv	Food industry portfolio
13	Beer	Kenneth French	lv	Beer industry portfolio
14	Smoke	Kenneth French	lv	Smoke industry portfolio
15	Toys	Kenneth French	lv	Toys industry portfolio
16	Fun	Kenneth French	lv	Fun industry portfolio
17	Books	Kenneth French	lv	Books industry portfolio
18	Hshld	Kenneth French	lv	Hshld industry portfolio
19	Clths	Kenneth French	lv	Clths industry portfolio
20	MedEq	Kenneth French	lv	MedEq industry portfolio
21	Drugs	Kenneth French	lv	Drugs industry portfolio
22	Chems	Kenneth French	lv	Chems industry portfolio
23	Rubbr	Kenneth French	lv	Rubbr industry portfolio
24	Txtls	Kenneth French	lv	Txtls industry portfolio
25	BldMt	Kenneth French	lv	BldMt industry portfolio
26	Cnstr	Kenneth French	lv	Cnstr industry portfolio
27	Steel	Kenneth French	lv	Steel industry portfolio
28	Mach	Kenneth French	lv	Mach industry portfolio
29	ElcEq	Kenneth French	lv	ElcEq industry portfolio
30	Autos	Kenneth French	lv	Autos industry portfolio
31	Aero	Kenneth French	lv	Aero industry portfolio
32	Ships	Kenneth French	lv	Ships industry portfolio
33	Mines	Kenneth French	lv	Mines industry portfolio
34	Coal	Kenneth French	lv	Coal industry portfolio
35	Oil	Kenneth French	lv	Oil industry portfolio
36	Util	Kenneth French	lv	Util industry portfolio
37	Telcm	Kenneth French	lv	Telcm industry portfolio
38	PerSv	Kenneth French	lv	PerSv industry portfolio
39	BusSv	Kenneth French	lv	BusSv industry portfolio
40	Hardw	Kenneth French	lv	Hardw industry portfolio
41	Chips	Kenneth French	lv	Chips industry portfolio
42	LabEq	Kenneth French	lv	LabEq industry portfolio
43	Paper	Kenneth French	lv	Paper industry portfolio
44	Boxes	Kenneth French	lv	Boxes industry portfolio
45	Trans	Kenneth French	lv	Trans industry portfolio
46	Whlsl	Kenneth French	lv	Whlsl industry portfolio
47	Rtail	Kenneth French	lv	Rtail industry portfolio
48	Meals	Kenneth French	lv	Meals industry portfolio
49	Banks	Kenneth French	lv	Banks industry portfolio
50	Insur	Kenneth French	lv	Insur industry portfolio
51	RlEst	Kenneth French	lv	RlEst industry portfolio
52	Fin	Kenneth French	lv	Fin industry portfolio
53	Other	Kenneth French	lv	Other industry portfolio
		Group 4: Size/BM		
54	1_2	Kenneth French	lv	(1, 2) portfolio sorted on (size, book-to-market)
55	1_4	Kenneth French	lv	(1, 4) portfolio sorted on (size, book-to-market)
56	1_5	Kenneth French	lv	(1, 5) portfolio sorted on (size, book-to-market)
57	1_6	Kenneth French	lv	(1, 6) portfolio sorted on (size, book-to-market)
58	1_7	Kenneth French	lv	(1, 7) portfolio sorted on (size, book-to-market)
59	1_8	Kenneth French	lv	(1, 8) portfolio sorted on (size, book-to-market)
60	1_9	Kenneth French	lv	(1, 9) portfolio sorted on (size, book-to-market)
61	1_high	Kenneth French	lv	(1, high) portfolio sorted on (size, book-to-market)
62	2_low	Kenneth French	lv	(2, low) portfolio sorted on (size, book-to-market)

Table A.3 (Cont'd)

No.	Short Name	Source	Tran	Description
63	2_2	Kenneth French	*lv*	(2, 2) portfolio sorted on (size, book-to-market)
64	2_3	Kenneth French	*lv*	(2, 3) portfolio sorted on (size, book-to-market)
65	2_4	Kenneth French	*lv*	(2, 4) portfolio sorted on (size, book-to-market)
66	2_5	Kenneth French	*lv*	(2, 5) portfolio sorted on (size, book-to-market)
67	2_6	Kenneth French	*lv*	(2, 6) portfolio sorted on (size, book-to-market)
68	2_7	Kenneth French	*lv*	(2, 7) portfolio sorted on (size, book-to-market)
69	2_8	Kenneth French	*lv*	(2, 8) portfolio sorted on (size, book-to-market)
70	2_9	Kenneth French	*lv*	(2, 9) portfolio sorted on (size, book-to-market)
71	2_high	Kenneth French	*lv*	(2, high) portfolio sorted on (size, book-to-market)
72	3_low	Kenneth French	*lv*	(3, low) portfolio sorted on (size, book-to-market)
73	3_2	Kenneth French	*lv*	(3, 2) portfolio sorted on (size, book-to-market)
74	3_3	Kenneth French	*lv*	(3, 3) portfolio sorted on (size, book-to-market)
75	3_4	Kenneth French	*lv*	(3, 4) portfolio sorted on (size, book-to-market)
76	3_5	Kenneth French	*lv*	(3, 5) portfolio sorted on (size, book-to-market)
77	3_6	Kenneth French	*lv*	(3, 6) portfolio sorted on (size, book-to-market)
78	3_7	Kenneth French	*lv*	(3, 7) portfolio sorted on (size, book-to-market)
79	3_8	Kenneth French	*lv*	(3, 8) portfolio sorted on (size, book-to-market)
80	3_9	Kenneth French	*lv*	(3, 9) portfolio sorted on (size, book-to-market)
81	3_high	Kenneth French	*lv*	(3, high) portfolio sorted on (size, book-to-market)
82	4_low	Kenneth French	*lv*	(4, low) portfolio sorted on (size, book-to-market)
83	4_2	Kenneth French	*lv*	(4, 2) portfolio sorted on (size, book-to-market)
84	4_3	Kenneth French	*lv*	(4, 3) portfolio sorted on (size, book-to-market)
85	4_4	Kenneth French	*lv*	(4, 4) portfolio sorted on (size, book-to-market)
86	4_5	Kenneth French	*lv*	(4, 5) portfolio sorted on (size, book-to-market)
87	4_6	Kenneth French	*lv*	(4, 6) portfolio sorted on (size, book-to-market)
88	4_7	Kenneth French	*lv*	(4, 7) portfolio sorted on (size, book-to-market)
89	4_8	Kenneth French	*lv*	(4, 8) portfolio sorted on (size, book-to-market)
90	4_9	Kenneth French	*lv*	(4, 9) portfolio sorted on (size, book-to-market)
91	4_high	Kenneth French	*lv*	(4, high) portfolio sorted on (size, book-to-market)
92	5_low	Kenneth French	*lv*	(5, low) portfolio sorted on (size, book-to-market)
93	5_2	Kenneth French	*lv*	(5, 2) portfolio sorted on (size, book-to-market)
94	5_3	Kenneth French	*lv*	(5, 3) portfolio sorted on (size, book-to-market)
95	5_4	Kenneth French	*lv*	(5, 4) portfolio sorted on (size, book-to-market)
96	5_5	Kenneth French	*lv*	(5, 5) portfolio sorted on (size, book-to-market)
97	5_6	Kenneth French	*lv*	(5, 6) portfolio sorted on (size, book-to-market)
98	5_7	Kenneth French	*lv*	(5, 7) portfolio sorted on (size, book-to-market)
99	5_8	Kenneth French	*lv*	(5, 8) portfolio sorted on (size, book-to-market)
100	5_9	Kenneth French	*lv*	(5, 9) portfolio sorted on (size, book-to-market)
101	5_high	Kenneth French	*lv*	(5, high) portfolio sorted on (size, book-to-market)
102	6_low	Kenneth French	*lv*	(6, low) portfolio sorted on (size, book-to-market)
103	6_2	Kenneth French	*lv*	(6, 2) portfolio sorted on (size, book-to-market)
104	6_3	Kenneth French	*lv*	(6, 3) portfolio sorted on (size, book-to-market)
105	6_4	Kenneth French	*lv*	(6, 4) portfolio sorted on (size, book-to-market)
106	6_5	Kenneth French	*lv*	(6, 5) portfolio sorted on (size, book-to-market)
107	6_6	Kenneth French	*lv*	(6, 6) portfolio sorted on (size, book-to-market)
108	6_7	Kenneth French	*lv*	(6, 7) portfolio sorted on (size, book-to-market)
109	6_8	Kenneth French	*lv*	(6, 8) portfolio sorted on (size, book-to-market)
110	6_9	Kenneth French	*lv*	(6, 9) portfolio sorted on (size, book-to-market)
111	6_high	Kenneth French	*lv*	(6, high) portfolio sorted on (size, book-to-market)
112	7_low	Kenneth French	*lv*	(7, low) portfolio sorted on (size, book-to-market)
113	7_2	Kenneth French	*lv*	(7, 2) portfolio sorted on (size, book-to-market)
114	7_3	Kenneth French	*lv*	(7, 3) portfolio sorted on (size, book-to-market)
115	7_4	Kenneth French	*lv*	(7, 4) portfolio sorted on (size, book-to-market)
116	7_5	Kenneth French	*lv*	(7, 5) portfolio sorted on (size, book-to-market)
117	7_6	Kenneth French	*lv*	(7, 6) portfolio sorted on (size, book-to-market)
118	7_7	Kenneth French	*lv*	(7, 7) portfolio sorted on (size, book-to-market)
119	7_8	Kenneth French	*lv*	(7, 8) portfolio sorted on (size, book-to-market)
120	7_9	Kenneth French	*lv*	(7, 9) portfolio sorted on (size, book-to-market)
121	8_low	Kenneth French	*lv*	(8, low) portfolio sorted on (size, book-to-market)
122	8_2	Kenneth French	*lv*	(8, 2) portfolio sorted on (size, book-to-market)
123	8_3	Kenneth French	*lv*	(8, 3) portfolio sorted on (size, book-to-market)
124	8_4	Kenneth French	*lv*	(8, 4) portfolio sorted on (size, book-to-market)
125	8_5	Kenneth French	*lv*	(8, 5) portfolio sorted on (size, book-to-market)
126	8_6	Kenneth French	*lv*	(8, 6) portfolio sorted on (size, book-to-market)
127	8_7	Kenneth French	*lv*	(8, 7) portfolio sorted on (size, book-to-market)
128	8_8	Kenneth French	*lv*	(8, 8) portfolio sorted on (size, book-to-market)

Table A.3 (Cont'd)

No.	Short Name	Source	Tran	Description
129	8_9	Kenneth French	*lv*	(8, 9) portfolio sorted on (size, book-to-market)
130	8_high	Kenneth French	*lv*	(8, high) portfolio sorted on (size, book-to-market)
131	9_low	Kenneth French	*lv*	(9, low) portfolio sorted on (size, book-to-market)
132	9_2	Kenneth French	*lv*	(9, 2) portfolio sorted on (size, book-to-market)
133	9_3	Kenneth French	*lv*	(9, 3) portfolio sorted on (size, book-to-market)
134	9_4	Kenneth French	*lv*	(9, 4) portfolio sorted on (size, book-to-market)
135	9_5	Kenneth French	*lv*	(9, 5) portfolio sorted on (size, book-to-market)
136	9_6	Kenneth French	*lv*	(9, 6) portfolio sorted on (size, book-to-market)
137	9_7	Kenneth French	*lv*	(9, 7) portfolio sorted on (size, book-to-market)
138	9_8	Kenneth French	*lv*	(9, 8) portfolio sorted on (size, book-to-market)
139	9_high	Kenneth French	*lv*	(9, high) portfolio sorted on (size, book-to-market)
140	10_low	Kenneth French	*lv*	(10, low) portfolio sorted on (size, book-to-market)
141	10_2	Kenneth French	*lv*	(10, 2) portfolio sorted on (size, book-to-market)
142	10_3	Kenneth French	*lv*	(10, 3) portfolio sorted on (size, book-to-market)
143	10_4	Kenneth French	*lv*	(10, 4) portfolio sorted on (size, book-to-market)
144	10_5	Kenneth French	*lv*	(10, 5) portfolio sorted on (size, book-to-market)
145	10_6	Kenneth French	*lv*	(10, 6) portfolio sorted on (size, book-to-market)
146	10_7	Kenneth French	*lv*	(10, 7) portfolio sorted on (size, book-to-market)
147	VXO	Fred MD	*lv*	VXOCLSx

CRSP Data Details Value-weighted price and dividend data were obtained from the Center for Research in Security Prices (CRSP). From the Annual Update data, we obtain monthly value-weighted returns series vwretd (with dividends) and vwretx (excluding dividends). These series have the interpretation

$$VWRETD_t = \frac{P_{t+1} + D_{t+1}}{P_t}$$
$$VWRETX_t = \frac{P_{t+1}}{P_t}$$

From these series, a normalized price series P, can be constructed using the recursion

$$P_0 = 1$$
$$P_t = P_{t-1} \cdot VWRETX_t.$$

A dividend series can then be constructed using

$$D_t = P_{t-1}(VWRETD_t - VWRETX_t).$$

In order to remove seasonality of dividend payments from the data, instead of D_t we use the series

$$D_t^* = \frac{1}{12} \sum_{j=0}^{11} D_{t-j}$$

i.e., the moving average over the entire year. For the price and dividend series under "reinvestment," we calculate the price under reinvestment, P_t^{re}, as the normalized value of the market portfolio under reinvestment of dividends, using the recursion

$$P_0^{re} = 1$$
$$P_t^{re} = P_{t-1} \cdot VWRETD_t$$

Similarly, we can define dividends under reinvestment, D_t^{re}, as the total dividend payments on this portfolio (the number of "shares" of which have increased over time) using

$$D_t^{re} = P_{t-1}^{re}(VWRETD_t - VWRETX_t).$$

As before, we can remove seasonality by using

$$D_t^{re,*} = \frac{1}{2}\sum_{j=0}^{11} D_{t-j}^{re}.$$

Five data series are constructed from the CRSP data as follows:

- D_log(DIV): $\Delta \log D_t^*$.
- D_log(P): $\Delta \log P_t$.
- D_DIVreinvest: $\Delta \log D_t^{re,*}$
- D_Preinvest: $\Delta \log P_t^{re,*}$
- d-p: $\log(D_t^*) - \log(P_t)$

Kenneth French Data Details The following data are obtained from the data library of Kenneth French's Dartmouth website (http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_libr

- Fama/French Factors: From this dataset we obtain the data series RF, Mkt-RF, SMB, HML.
- 25 Portfolios formed on Size and Book-to-Market (5 x 5): From this dataset we obtain the series R15-R11, which is the spread between the (small, high book-to-market) and (small, low book-to-market) portfolios.
- Momentum Factor (Mom): From this dataset we obtain the series UMD, which is equal to the momentum factor.
- 49 Industry Portfolios: From this dataset we use all value-weighted series, excluding any series that have missing observations from Jan. 1960 on, from which we obtain the series Agric through Other. The omitted series are: Soda, Hlth, FabPr, Guns, Gold, Softw.
- 100 Portfolios formed in Size and Book-to-Market: From this dataset we use all value-weighted series, excluding any series that have missing observations from Jan. 1960 on. This yields variables with the name X_Y where X stands for the index of the size variable (1, 2, ..., 10) and Y stands for the index of the book-to-market variable (Low, 2, 3, ..., 8, 9, High). The omitted series are 1_low, 1_3, 7_high, 9_9, 10_8, 10_9, 10_high.

5.0.3 Daily Financial Data

Daily Data and construction of daily factors The daily financial series in this data set are from the daily financial dataset used in Andreou, Ghysels, and Kourtellos (2013). We create a smaller daily database which is a subset of the large cross-section of 991 daily series in their dataset. Our dataset covers five classes of financial assets: (i) the Commodities class; (ii) the Corporate Risk category; (iii) the Equities class; (iv) the Foreign Exchange Rates class and (v) the Government Securities.

The dataset includes up to 87 daily predictors in a daily frequency from 23-Oct-1959 to 24-Oct-2018 (14852 trading days) from the above five categories of financial assets. We remove series with fewer than ten years of data and time periods with no variables observed, which occurs for some series in the early part of the sample. For those years, we have less than 87 series. There are 39 commodity variables which include commodity indices, prices and futures, 16 corporate risk series, 9 equity series which include major US stock market indices and the 500 Implied Volatility, 16 government securities which include the federal funds rate, government treasury bills of securities from three months to ten years, and 7 foreign exchange variables which include the individual foreign exchange rates of major five US trading partners and two effective exchange rate. We choose these daily predictors because they are proposed in the literature as good predictors of economic growth.

We construct daily financial factors in a quarterly frequency in two steps. First, we use these daily financial time series to form factors at a daily frequency. The raw data used to form factors are always transformed to achieve stationarity. The raw daily data are also standardized before performing factor estimation (see generic description below). We estimate factors at each daily date in the sample using the entire history (from 23-Oct-1959) of variables observed in real time.

In the second step, we convert these daily financial indicators to quarterly weighted variables to form quarterly factors using the optimal weighting scheme according to the method described below (see the optimal weighting scheme section).

The data series used in this dataset are listed below in Table A.4 by data source. The tables also list the transformation applied to each variable to make them stationary before generating factors. The transformations used to stationarize a time series are the same as those explained in the section "Monthly financial factor data".

Table A.4: List of Daily Financial Dataset Variables

No.	Short Name	Source	Tran	Description
		Group 1: Commodities		
1	GSIZSPT	Data Stream	Δln	S&P GSCI Zinc Spot - PRICE INDEX
2	GSSBSPT	Data Stream	Δln	S&P GSCI Sugar Spot - PRICE INDEX
3	GSSOSPT	Data Stream	Δln	S&P GSCI Soybeans Spot - PRICE INDEX
4	GSSISPT	Data Stream	Δln	S&P GSCI Silver Spot - PRICE INDEX
5	GSIKSPT	Data Stream	Δln	S&P GSCI Nickel Spot - PRICE INDEX
6	GSLCSPT	Data Stream	Δln	S&P GSCI Live Cattle Spot - PRICE INDEX
7	GSLHSPT	Data Stream	Δln	S&P GSCI Lean Hogs Index Spot - PRICE INDEX
8	GSILSPT	Data Stream	Δln	S&P GSCI Lead Spot - PRICE INDEX
9	GSGCSPT	Data Stream	Δln	S&P GSCI Gold Spot - PRICE INDEX
10	GSCTSPT	Data Stream	Δln	S&P GSCI Cotton Spot - PRICE INDEX
11	GSKCSPT	Data Stream	Δln	S&P GSCI Coffee Spot - PRICE INDEX
12	GSCCSPT	Data Stream	Δln	S&P GSCI Cocoa Index Spot - PRICE INDEX

Table A.4 (Cont'd)

No.	Short Name	Source	Tran	Description
13	GSIASPT	Data Stream	Δln	S&P GSCI Aluminum Spot - PRICE INDEX
14	SGWTSPT	Data Stream	Δln	S&P GSCI All Wheat Spot - PRICE INDEX
15	EIAEBRT	Data Stream	Δln	Europe Brent Spot FOB U$/BBL Daily
16	CRUDOIL	Data Stream	Δln	Crude Oil-WTI Spot Cushing U$/BBL - MID PRICE
17	LTICASH	Data Stream	Δln	LME-Tin 99.85% Cash U$/MT
18	CWFCS00	Data Stream	Δln	CBT-WHEAT COMPOSITE FUTURES CONT. - SETT. PRICE
19	CCFCS00	Data Stream	Δln	CBT-CORN COMP. CONTINUOUS - SETT. PRICE
20	CSYCS00	Data Stream	Δln	CBT-SOYBEANS COMP. CONT. - SETT. PRICE
21	NCTCS20	Data Stream	Δln	CSCE-COTTON #2 CONT.2ND FUT - SETT. PRICE
22	NSBCS00	Data Stream	Δln	CSCE-SUGAR #11 CONTINUOUS - SETT. PRICE
23	NKCCS00	Data Stream	Δln	CSCE-COFFEE C CONTINUOUS - SETT. PRICE
24	NCCCS00	Data Stream	Δln	CSCE-COCOA CONTINUOUS - SETT. PRICE
25	CZLCS00	Data Stream	Δln	ECBOT-SOYBEAN OIL CONTINUOUS - SETT. PRICE
26	COFC01	Data Stream	Δln	CBT-OATS COMP. TRc1 - SETT. PRICE
27	CLDCS00	Data Stream	Δln	CME-LIVE CATTLE COMP. CONTINUOUS - SETT. PRICE
28	CLGC01	Data Stream	Δln	CME-LEAN HOGS COMP. TRc1 - SETT. PRICE
29	NGCCS00	Data Stream	Δln	CMX-GOLD 100 OZ CONTINUOUS - SETT. PRICE
30	LAH3MTH	Data Stream	Δln	LME-Aluminium 99.7% 3 Months U$/MT
31	LED3MTH	Data Stream	Δln	LME-Lead 3 Months U$/MT
32	LNI3MTH	Data Stream	Δln	LME-Nickel 3 Months U$/MT
33	LTI3MTH	Data Stream	Δln	LME-Tin 99.85% 3 Months U$/MT
34	PLNYD	www.macrotrends.net	Δln	Platinum Cash Price (U$ per troy ounce)
35	XPDD	www.macrotrends.net	Δln	Palladium (U$ per troy ounce)
36	CUS2D	www.macrotrends.net	Δln	Corn Spot Price (U$/Bushel)
37	SoybOil	www.macrotrends.net	Δln	Soybean Oil Price (U$/Pound)
38	OATSD	www.macrotrends.net	Δln	Oat Spot Price (US$/Bushel)
39	WTIOilFut	US EIA	Δln	Light Sweet Crude Oil Futures Price: 1St Expiring Contract Settlement ($/Bbl)
		Group 2: Equities		
40	S&PCOMP	Data Stream	Δln	S&P 500 COMPOSITE - PRICE INDEX
41	ISPCS00	Data Stream	Δln	CME-S&P 500 INDEX CONTINUOUS - SETT. PRICE
42	SP5EIND	Data Stream	Δln	S&P500 ES INDUSTRIALS - PRICE INDEX
43	DJINDUS	Data Stream	Δln	DOW JONES INDUSTRIALS - PRICE INDEX
44	CYMCS00	Data Stream	Δln	CBT-MINI DOW JONES CONTINUOUS - SETT. PRICE
45	NASCOMP	Data Stream	Δln	NASDAQ COMPOSITE - PRICE INDEX
46	NASA100	Data Stream	Δln	NASDAQ 100 - PRICE INDEX
47	CBOEVIX	Data Stream	lv	CBOE SPX VOLATILITY VIX (NEW) - PRICE INDEX
48	S&P500toVIX	Data Stream	Δln	S&P500/VIX
		Group 3: Corporate Risk		
49	LIBOR	FRED	Δlv	Overnight London Interbank Offered Rate (%)
50	1MLIBOR	FRED	Δlv	1-Month London Interbank Offered Rate (%)
51	3MLIBOR	FRED	Δlv	3-Month London Interbank Offered Rate (%)
52	6MLIBOR	FRED	Δlv	6-Month London Interbank Offered Rate (%)
53	1YLIBOR	FRED	Δlv	One-Year London Interbank Offered Rate (%)
54	1MEuro-FF	FRED	lv	1-Month Eurodollar Deposits (London Bid) (% P.A.) minus Fed Funds
55	3MEuro-FF	FRED	lv	3-Month Eurodollar Deposits (London Bid) (% P.A.) minus Fed Funds
56	6MEuro-FF	FRED	lv	6-Month Eurodollar Deposits (London Bid) (% P.A.) minus Fed Funds
57	APFNF-AANF	Data Stream	lv	1-Month A2/P2/F2 Nonfinancial Commercial Paper (NCP) (% P. A.) minus 1-Month Aa NCP (% P.A.)
58	APFNF-AAF	Data Stream	lv	1-Month A2/P2/F2 NCP (% P.A.) minus 1-Month Aa Financial Commercial Paper (% P.A.)
59	TED	Data Stream, FRED	lv	3Month Tbill minus 3-Month London Interbank Offered Rate (%)
60	MAaa-10YTB	Data Stream	lv	Moody Seasoned Aaa Corporate Bond Yield (% P.A.) minus Y10-Tbond
61	MBaa-10YTB	Data Stream	lv	Moody Seasoned Baa Corporate Bond Yield (% P.A.) minus Y10-Tbond
62	MLA-10YTB	Data Stream, FRED	lv	Merrill Lynch Corporate Bonds: A Rated: Effective Yield (%) minus Y10-Tbond
63	MLAA-10YTB	Data Stream, FRED	lv	Merrill Lynch Corporate Bonds: Aa Rated: Effective Yield (%) minus Y10-Tbond

Table A.4 (Cont'd)

No.	Short Name	Source	Tran	Description
64	MLAAA-10YTB	Data Stream, FRED	lv	Merrill Lynch Corporate Bonds: Aaa Rated: Effective Yield (%) minus Y10-Tbond
		Group 4: Treasuries		
65	FRFEDFD	Data Stream	Δlv	US FED FUNDS EFF RATE (D) - MIDDLE RATE
66	FRTBS3M	Data Stream	Δlv	US T-BILL SEC MARKET 3 MONTH (D) - MIDDLE RATE
67	FRTBS6M	Data Stream	Δlv	US T-BILL SEC MARKET 6 MONTH (D) - MIDDLE RATE
68	FRTCM1Y	Data Stream	Δlv	US TREASURY CONST MAT 1 YEAR (D) - MIDDLE RATE
69	FRTCM10	Data Stream	Δlv	US TREASURY CONST MAT 10 YEAR (D) - MIDDLE RATE
70	6MTB-FF	Data Stream	lv	6-month treasury bill market bid yield at constant maturity (%) minus Fed Funds
71	1YTB-FF	Data Stream	lv	1-year treasury bill yield at constant maturity (% P.A.) minus Fed Funds
72	10YTB-FF	Data Stream	lv	10-year treasury bond yield at constant maturity (% P.A.) minus Fed Funds
73	6MTB-3MTB	Data Stream	lv	6-month treasury bill yield at constant maturity (% P.A.) minus 3M-Tbills
74	1YTB-3MTB	Data Stream	lv	1-year treasury bill yield at constant maturity (% P.A.) minus 3M-Tbills
75	10YTB-3MTB	Data Stream	lv	10-year treasury bond yield at constant maturity (% P.A.) minus 3M-Tbills
76	BKEVEN05	FRB	lv	US Inflation compensation: continuously compounded zero-coupon yield: 5-year (%)
77	BKEVEN10	FRB	lv	US Inflation compensation: continuously compounded zero-coupon yield: 10-year (%)
78	BKEVEN1F4	FRB	lv	BKEVEN1F4
79	BKEVEN1F9	FRB	lv	BKEVEN1F9
80	BKEVEN5F5	FRB	lv	US Inflation compensation: coupon equivalent forward rate: 5-10 years (%)
		Group 5: Foreign Exchange (FX)		
81	US_CWBN	Data Stream	Δln	US NOMINAL DOLLAR BROAD INDEX - EXCHANGE INDEX
82	US_CWMN	Data Stream	Δln	US NOMINAL DOLLAR MAJOR CURR INDEX - EXCHANGE INDEX
83	US_CSFR2	Data Stream	Δln	CANADIAN $ TO US $ NOON NY - EXCHANGE RATE
84	EU_USFR2	Data Stream	Δln	EURO TO US$ NOON NY - EXCHANGE RATE
85	US_YFR2	Data Stream	Δln	JAPANESE YEN TO US $ NOON NY - EXCHANGE RATE
86	US_SFFR2	Data Stream	Δln	SWISS FRANC TO US $ NOON NY - EXCHANGE RATE
87	US_UKFR2	Data Stream	Δln	UK POUND TO US $ NOON NY - EXCHANGE RATE

From Daily to Quarterly Factors: Weighting Schemes After we obtain daily financial factors $\mathbf{G}_{D,t}$, we use some weighting schemes proposed in the literature about Mixed Data Sampling (MIDAS) regressions to form quarterly factors, $\mathbf{G}_{D,t}^{Q}$. Denote by G_t^D a factor in a daily frequency formed from the daily financial dataset and denote by G_t^Q a quarterly aggregate of the corresponding daily factor time series. Let $G_{N_D-j,d_t,t}^D$ denote the value of a daily factor in the j^{th} day counting backwards from the survey deadline d_t in quarter t. Hence, the day d_t of quarter t corresponds with $j=0$ and is therefore $G_{N_D,d_t,t}^D$. For simplicity, we suppress the subscript d_t thus $G_{N_D-j,d_t,t}^D \equiv G_{N_D-j,t}^D$.

We compute the quarterly aggregate of a daily financial factor as a weighted average of observations over the N_D business days before the survey deadline. This means that the forecasters's information set includes daily financial data up to the previous N_D business days. G_t^Q is defined as:

$$G_t^Q(\mathbf{w}) \equiv \sum_{i=1}^{N_D} w_i G_{N_D-i,t}^D$$

where $\mathbf{w}$ is a vector of weights. We consider the following three types of weighting schemes to convert daily factor observations to quarterly. Each weighting scheme weights information by some function of the number of days prior to the survey deadline.

1. $w_i = 1$ for $i = 1$ and $w_i = 0$ otherwise. This weighting scheme places all weight on data in the last business day before the survey deadline for that quarter and zero weight on any data prior to that day.

2. $w_i = \frac{\theta^j}{\sum_{j=1}^{N_D} \theta^j}$ where we consider a range of θ^j for $\theta^j = (0.1, 0.2, 0.3, 0.7, 0.8, 0.9, 1)'$. The smaller is θ^j, the more rapidly information prior to the survey deadline day is downweighted. This down-weighting is progressive but not nonmonotone. $\theta^j = 1$ is a simple average of the observations across all days in the quarter.

3. The third parameterization has two parameters, or $\theta^D = (\theta_1, \theta_2)'$ and allows for non-monotone weighting of past information:

$$w\left(i; \theta_1, \theta_2\right) = \frac{f\left(\frac{i}{N_D}, \theta_1; \theta_2\right)}{\sum_{j=1}^{N_D} f\left(\frac{j}{N_D}, \theta_1; \theta_2\right)}$$

where:

$$f(x, a, b) = \frac{x^{a-1}(1-x)^{b-1}\Gamma(a+b)}{\Gamma(a)\Gamma(b)}$$

$$\Gamma(a) = \int_0^\infty e^{-x} x^{a-1} dx$$

The weights $w\left(i; \theta_1, \theta_2\right)$ are the Beta polynomial MIDAS weights of Ghysels, Sinko, and Valkanov (2007), which are based on the Beta function. This weighting scheme is flexible enough to generate a range of possible shapes with only two parameters.

We consider these possible weighting schemes and choose the optimal weighting scheme $\mathbf{w}^*$ from these 24 weighting schemes for a daily financial factor G_t^D by minimizing the sum of square residuals in a regression of $y_{j,t+h}$ on $G_t^Q(\mathbf{w})$:

$$y_{j,t+h} = a + b \cdot \underbrace{\sum_{i=1}^{N_D} w_i G_{N_D - i, t}^D}_{G_t^Q(\mathbf{w})} + u_{t+h}.$$

This is done in real time using recursive regressions and an initial in-sample estimation window that matches the timing described below for the data-dependent choice of tuning parameter in the machine learning estimation (see the section on Estimation and Machine Learning).

We assume that $N_D = 14$ which implies that forecasters use daily information in at most the past two weeks before the survey deadline. The process is repeated for each daily financial factor in $\mathbf{G}_{D,t}$ to form quarterly factors $\mathbf{G}_{D,t}^Q$.

Estimation and Machine Learning

The model to be estimated is

$$y_{j,t+h} = \mathcal{X}_t' \boldsymbol{\beta}_j^{(i)} + \epsilon_{jt+h}.$$

It should be noted that the most recent observation on the left-hand-side is generally available in real time only with a one-period lag, thus the forecasting estimations can only be run with data over a sample that stops one period later than today in real time. $\mathcal{X}_t$ always denotes the most recent data that would have been in real time prior to the date on which the forecast was submitted.

The coefficients $\boldsymbol{\beta}_{j,t}^{(i)}$ are estimated using the Elastic Net (EN) estimator, which depend on regularization parameter parameters $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)'$ (See the next section for a description of EN).

The procedure involves iterating on the following steps.

1. **Sample partitioning**: At time t, a prior sample of size $\widetilde{T}$ is partitioned into two subsample windows: an "in-sample" estimation subsample consisting of the first T_{IS} observations, and a hold-out "training" subsample of T_{TS} subsequent observations, i.e., $\widetilde{T} = T_{IS} + T_{TS}$.

2. **In-sample estimation**: Initial estimates of $\boldsymbol{\beta}^{(i)}$ are obtained using the EN estimator using observations $1, ..., T_{IS}$, given an arbitrary fixed (non-random) starting value for $\boldsymbol{\lambda}$. Denote this initial estimate $\boldsymbol{\beta}_{T_{IS}}^{*(i)}(\mathbf{X}_{T_{IS}}, \boldsymbol{\lambda})$, where "*" denotes the value of the estimator given an arbitrary $\boldsymbol{\lambda}$.

3. **Training and cross-validation**: The regularization parameter $\boldsymbol{\lambda}$ is estimated by minimizing mean-square loss $\mathcal{L}(\boldsymbol{\lambda}, T_{IS}, T_{TS})$ over pseudo-out-of-sample forecast errors generated from rolling regressions using only the most recent T_{IS} observations. That is, the first rolling prediction uses data from 1 to T_{IS}, the second rolling prediction uses data from 2 to $T_{IS} + 1$, etc., where

$$\mathcal{L}(\boldsymbol{\lambda}, T_{IS}, T_{TS}) \equiv \frac{1}{T_{TS} - h} \sum_{\tau=T_{IS}}^{T_{IS}+T_{TS}-h} \left(\mathcal{X}_\tau' \boldsymbol{\beta}_{j,\tau}^{*(i)}(\mathbf{X}_{T_{IS}}, \boldsymbol{\lambda}) - y_{j,\tau+h} \right)^2,$$

and where $\boldsymbol{\beta}_{j,\tau}^{*(i)}(\mathbf{X}_{T_{IS}}, \boldsymbol{\lambda})$ is the time τ EN estimate of $\boldsymbol{\beta}_j^{(i)}$ given $\boldsymbol{\lambda}$ and data through time τ in a sample of size T_{IS}.

4. Steps 1-3 are repeated over a grid of estimation and training sample window lengths T_{IS}^* and T_{TS}^* such that alternative partitions satisfy $T_{IS}^* + T_{TS}^* \leq \widetilde{T}$, where shorter window lengths remove consecutive observations at the start of the prior sample. The final machine estimate of $\boldsymbol{\beta}_{j,t}^{(i)}(\mathbf{X}_{\widetilde{T}}, \boldsymbol{\lambda})$ uses $\left\{ \widehat{\boldsymbol{\lambda}}, \widehat{T}_{IS}, \widehat{T}_{IS} \right\} = \underset{\boldsymbol{\lambda}, T_{IS}^*, T_{TS}^*}{\operatorname{argmin}} \mathcal{L}(\boldsymbol{\lambda}, T_{IS}^*, T_{TS}^*)$ and is denoted $\widehat{\boldsymbol{\beta}}_{j,t}^{(i)}\left(\mathbf{X}_{\widetilde{T}}, \widehat{\boldsymbol{\lambda}}\right)$.

5. **Out-of-sample prediction**: The values of the regressors at time t are used to make a true out-of-sample prediction of y_{t+h}, using $\widehat{\boldsymbol{\beta}}_{j,t}^{(i)}\left(\mathbf{X}_{\widetilde{T}}, \widehat{\boldsymbol{\lambda}}\right)$, and the machine forecast error $y_{t+h} - \mathcal{X}_t' \widehat{\boldsymbol{\beta}}_{j,t}^{(i)}\left(\mathbf{X}_{\widetilde{T}}, \widehat{\boldsymbol{\lambda}}\right)$ stored.

6. **Roll forward and repeat**: The initial in-sample subperiod is rolled forward one period and uses data from $2, ..., T_{IS} + 1$ and steps 2-5 are repeated until the final out-of-sample forecast is made for $y_{j,T}$, where T is the last period of our sample.

Averaging across the forecast errors from step 5 above, gives $\text{MSE}_{\mathbb{E}}$ over the *evaluation sample* $t = (T_{IS} + T_{TS} + h)$,...,$T$ that we use to assess whether belief distortions are present.

Any such distortions are quantified by looking at the ratio $\text{MSE}_{\mathbb{E}}/\text{MSE}_{\mathbb{F}}$ over the evaluation sample.

The procedure can be understood for a specific example. Let $h = 4$ quarters. The initial in-sample period is 1969:Q1-1973:Q4, corresponding to $T_{IS} = 20$. Suppose $T_{TS} = 16$. The initial in-sample regression is run for $t =$1969:Q1-1973:Q4 (dependent variable from 1970:Q1-1973:Q4, independent variable from 1969:Q1-1972:Q4), and the values of the regressors at $t =$ 1973:Q4 are used to forecast y_j for 1973:Q4-1974:Q4, which serves as the first training forecast. All parameters are then reestimated from 1969:Q2-1974:Q1 and forecasts are recomputed for y_j for 1974:Q1-1975:Q1 and so on, until the final training forecast is made for y_j for 1977:Q3 to 1978:Q3. The squared forecast errors from the 16 training forecasts are stored and the regularization parameters $\boldsymbol{\lambda}$ are chosen to minimize the mean-square-forecast error over the training sample, with this chosen value denoted $\boldsymbol{\lambda}^*$. Next, the regressors at $t =$1978:Q3 are multiplied by $\hat{\boldsymbol{\beta}}_j^{(\tau)}(\boldsymbol{\lambda}^*)$ to produce an out-of-sample forecast of y_j for 1978:Q3-1979:Q3, which is the first observation in our "evaluation" sample. We repeat this on a grid of (T_{IS},T_{TS}), and the optimal combination of T_{IS} and T_{TS} minimizes $MSE_{j\mathbb{E}}(\boldsymbol{\lambda}^*, T_{IS}, T_{TS})$. The entire procedure is repeated by rolling forward to the next in-sample period 1969:Q1-1974:Q1.

We allow the machine to additionally learn about whether the coefficient on the survey forecast should be shrunk toward zero or toward unity. Recall that The machine forecast for the ith percentile is

$$\mathbb{E}_t^{(i)}(y_{j,t+h}) \equiv \hat{\alpha}_j^{(i)} + \hat{\beta}_{j\mathbb{F}}^{(i)}\mathbb{F}_t^{(i)}[y_{j,t+h}] + \hat{\mathbf{B}}_{j\mathcal{Z}}^{(i)\prime}\mathcal{Z}_{jt}.$$

One possibility is to run the machine model as a regression of forecast errors on time t information:

$$y_{j,t+h} - \mathbb{F}_t^{(i)}[y_{j,t+h}] = \alpha_j^{(i)} + \beta_{j\mathbb{F}}^{(i)}\mathbb{F}_t^{(i)}[y_{j,t+h}] + \mathbf{B}_{j\mathcal{Z}}^{(i)\prime}\mathcal{Z}_t + \epsilon_{jt+h}, \tag{A.12}$$

where for this case the machine efficient benchmark is characterized by $\beta_{j\mathbb{F}}^{(i)} = 0$; $\mathbf{B}_{j\mathcal{Z}}^{(i)} = \mathbf{0}$;$\alpha_j^{(i)} = 0$. Because elastic net shrinks estimated coefficients toward zero, this centering of the left-hand-side variable allows the machine to shrink $\beta_{j\mathbb{F}}^{(i)}$ toward one. In this case the machine forecast is given by

$$\mathbb{E}_t^{(i)}(y_{j,t+h}) \equiv \hat{\alpha}_j^{(i)} + \left(\hat{\beta}_{j\mathbb{F}}^{(i)} + 1\right)\mathbb{F}_t^{(i)}[y_{j,t+h}] + \hat{\mathbf{B}}_{j\mathcal{Z}}^{(i)\prime}\mathcal{Z}_{jt}.$$

By contrast, if the machine forecast is implemented by running

$$y_{j,t+h} = \alpha_j^{(i)} + \beta_{j\mathbb{F}}^{(i)}\mathbb{F}_t^{(i)}[y_{j,t+h}] + \mathbf{B}_{j\mathcal{Z}}^{(i)\prime}\mathcal{Z}_t + \epsilon_{jt+h},$$

then $\beta_{j\mathbb{F}}^{(i)}$ is shrunk toward zero and the algorithm will typically place less weight on the survey forecast than if the estimation (A.12) is run. In the implementation, we allow the machine to choose which specification to run over time by having it pick the one that that minimizes the mean-square loss function $\mathcal{L}(\boldsymbol{\lambda}, T_{IS}, T_{TS})$ in every training sample.

To capture non-linearities, the machine forecasts follow a simple switching model. In most periods, the forecast is based on the "normal-times" statistical model just described. To cope with rapid economic change, as in a recession, the machine forecast is permitted to switch to a simpler model based on a recession indicator. We use as the recession indicator the term spread, defined as the difference between the 10-year Treasury bond rate and the 3-month Treasury bill rate. When the term spread at time t is at or below the real time sample 10th percentile value, the machine forecast of $t+4$ is switched to a recession-model forecast which is based solely on a dummy indicator I_{t-4}, which takes the value 1 when the term spread at $t-4$

is below a threshold. The precise threshold used is the one that minimizes the mean-square loss function in the relevant training sample prior to the actual forecast. The machine chooses among thresholds that represent the real time sample 10th, 5th, or 1st percentile values for the term spread. The recession model forecast is the fitted value from a regression of real time real GDP growth at time t on the 4-quarter lagged value of I_{t-4}.

Elastic Net

We use the Elastic Net (EN) estimator, which combines Least Absolute Shrinkage and Selection Operator (LASSO) and ridge type penalties. LASSO. Suppose our goal is to estimate the coefficients in the linear model:

$$y_{j,t+h} = \alpha_j + \beta_{j\mathbb{F}}\mathbb{F}_t^{(i)}\left[y_{j,t+h}\right] + \underbrace{\mathbf{B}_{j\mathcal{Z}}}_{qr\times qr}\mathcal{Z}_{jt} + \epsilon_{jt+h}$$

Collecting all the independent variables and coefficients into a single matrix and vector, the model can be written as:

$$y_{j,t+h} = \mathcal{X}_{tj}'\boldsymbol{\beta}_j + \epsilon_{jt+h}$$

where $\mathcal{X}_t = (1, \mathcal{X}_{1t,...,}\mathcal{X}_{Kt})'$ collects all the independent variable observations $\left(\mathbb{F}_t^{(i)}\left[y_{j,t+h}\right], \mathcal{Z}_{jt}\right)$ into a vector with "1" and $\boldsymbol{\beta}_j = \left(\alpha_j, \beta_{j\mathbb{F}}, \text{vec}\left(\mathbf{B}_{j\mathcal{Z}}\right)\right)' \equiv \left(\beta_0, \beta_1, ...\beta_K\right)'$ collects all the coefficient. It is customary to standardize the elements of $\mathcal{X}_t$ such that sample means are zero and sample standard deviations are unity. The coefficient estimates are then put back in their original scale by multiplying the slope coefficients by their respective standard deviations, and adding back the mean (scaled by slope coefficient over standard deviation.)

The EN estimator incorporates both an L_1 and L_2 penalty:

$$\hat{\boldsymbol{\beta}}^{\text{EN}} = \underset{\beta_0,\beta_1,...,\beta_k}{\text{argmin}}\left\{\sum_{t=1}^{T}\left(y_{j,t+h} - \mathcal{X}_t'\boldsymbol{\beta}_j\right)^2 + \underbrace{\lambda_1\sum_{j=1}^{k}\left|\boldsymbol{\beta}_j\right|}_{\text{LASSO}} + \underbrace{\lambda_2\sum_{j=1}^{k}\boldsymbol{\beta}_j^2}_{\text{ridge}}\right\}$$

By minimizing the MSE over the training samples, we choose the optimal λ_1 and λ_2 values simultaneously.

Dynamic Factor Estimation

We re-estimate factors at each date in the sample using the entire history of variables observed in real time. Let x_{it} denote the ith variable in a large dataset. The following steps are taken in forming the macro, financial, and daily factors:

1. Remove outlier values from a series, defined as values whose distance from the median is greater than ten times the interquartile range.

2. Scale each series according to the procedure proposed by Huang, Jiang, and Tong (2017). We run the following regression for each variable x_{it}:

$$y_{jt+h} = \beta_{j,i,0} + \beta_{j,i,x} x_{it} + \nu_{j,i,t+h}.$$

Then, we form a new dataset of variables $\hat{\beta}_{j,i,x} x_{it}$ where $\hat{\beta}_{j,i,x}$ denotes the OLS estimate of $\beta_{j,i,x}$. These "scaled" variables are standardized and denoted $\widetilde{x}_{it}$.

3. Throughout, the factors are estimated over $\widetilde{x}_{it}$ by the method of static principal components (PCA). The approach we consider is to posit that $\widetilde{x}_{it}$ has a factor structure taking the form

$$\widetilde{x}_{it} = \lambda_i' \mathbf{G}_t + e_{it}, \tag{A.13}$$

where $\mathbf{G}_t$ is a $r \times 1$ vector of latent common factors, λ_i is a corresponding $r \times 1$ vector of latent factor loadings, and e_{it} is a vector of idiosyncratic errors.[12] Specifically, the $T \times r$ matrix $\hat{g}_t$ is $\sqrt{T}$ times the r eigenvectors corresponding to the r largest eigenvalues of the $T \times T$ matrix $\widetilde{x}\widetilde{x}'/(TN_{\widetilde{x}})$ in decreasing order, where T is the number of time periods and $N_{\widetilde{x}}$ is the number of variables in the large dataset. The optimal number of common factors, r is determined by the panel information criteria developed in Bai and Ng (2002). To handle missing values in any series, we use an expectation-maximization (EM) algorithm by filling with an initial guess and forming factors, using (A.13) to update the guess with $\mathbb{E}(\widetilde{x}_{it}) = \mathbb{E}\left(\lambda_i' \hat{g}_t\right)$, and iterating until the successive values for $\mathbb{E}(\widetilde{x}_{it})$ are arbitrarily close.

4. Collect the common factors into the matrix $\mathbf{G}_{raw}$, where each principle component is a column.

5. Square the raw variables and repeat steps 2 through 5. Collect the common factors from squared data into a matrix $\mathbf{G}_{sqr}$, where component is a column.

6. Square the first factor in $\mathbf{G}_{raw}$, and call this $\mathbf{G}^2_{raw1}$.

7. Our matrix of factors is $[\mathbf{G}_{raw}, \mathbf{G}_{sqr1}, \mathbf{G}^2_{raw1}]$, where $\mathbf{G}_{sqr1}$ is the first column of $\mathbf{G}_{sqr}$.

For macro factors, we use all of the variables listed in Table A.2. After step 1 above, an additional step of removing missing variables and observations is needed for the macro variables. We remove series with fewer than seven years of data and time periods with less than fifty-percent of variables observed, which occur in the early part of the sample. Furthermore, we lag variables with missing data in the final observation whenever more than twenty-percent of variables are missing data in the last observation.[13]

For the financial factors, we use all of the variables listed in Table A.3, and no additional steps are performed beyond those described above.

[12] We consider an *approximate* dynamic factor structure, in which the idiosyncratic errors e_{it} are permitted to have a limited amount of cross-sectional correlation. The approximate factor specification limits the contribution of the idiosyncratic covariances to the total variance of x as N gets large:

$$N^{-1} \sum_{i=1}^{N} \sum_{j=1}^{N} \left| E\left(e_{it} e_{jt}\right) \right| \leq M,$$

where M is a constant.

[13] Even though the EM algorithm is designed to estimate missing observations, it does not perform well when there are too many missing observations at a single point in time.

Economic Interpretation of the Factors

Any labeling of the factors is imperfect because each is influenced to some degree by all the variables in the large dataset and the orthogonalization means that no one of them will correspond exactly to a precise economic concept like output or unemployment. Following Ludvigson and Ng (2009), we relate the factors to the underlying variables in the large dataset. For each time period in our evaluation sample, we compute the marginal R^2 from regressions of each of the individual series in the panel dataset onto each factor, one at a time. Each series $\widetilde{x}_{it}$ is assigned the group name in the data appendix tables naming all series, e.g., non-farm payrolls are part of the Employment group (EMP). If series $\widetilde{x}_{it}$ has the highest average marginal R^2 over all evaluation periods for factor G_{kt}, we label G_{kt} according to the group to which $\widetilde{x}_{it}$ belongs, e.g., G_{kt} is an Employment factor. We further normalize the sign of each factor so that an increase in the factor indicates an increase in $\widetilde{x}_{it}$. Thus, in the example above, an increase in G_{kt} would indicate a rise in non-farm payrolls. Table A.5 reports the series with largest average marginal R^2 for each factor of each large dataset.

Table A.5: Economic Interpretation of the Factors

	Series with Largest R^2	
	Macro Factors	Label
$G_{1,M,t}$	Nonfarm Payrolls	Macro Factor: Employment
$G_{2,M,t}$	Interest paid by consumers	Macro Factor: Money and Credit
$G_{3,M,t}$	Agg. Weekly hours - Service-producing	Macro Factor: Employment.
$G_{4,M,t}$	Agg. Weekly hours - Good-producing	Macro Factor: Employment
$G_{5,M,t}$	Nonborrowed Reserves	Macro Factor: Money and Credit
$G_{6,M,t}$	Housing Starts	Macro Factor: Orders and Investment
$G_{7,M,t}$	Change in private inventories	Macro Factor: Orders and Investment
$G_{8,M,t}$	PCE: Service	Macro Factor: Consumption
	Financial Factors	
$G_{1,F,t}$	D_log(P)	Financial Factor: Prices, Yield, Dividends
$G_{2,F,t}$	SMB	Financial Factor: Equity Risk Factors
$G_{3,F,t}$	HML	Financial Factor: Equity Risk Factors
$G_{4,F,t}$	R15_R11	Financial Factor: Equity Risk Factors
$G_{5,F,t}$	D_DIVreinvest	Financial Factor: Prices, Yield, Dividends
$G_{6,F,t}$	Smoke	Financial Factor: Industries
$G_{7,F,t}$	UMD	Financial Factor: Equity Risk Factors
$G_{8,F,t}$	Telcm	Financial Factor: Industries
	Daily Factors	
$G_{1,D,t}$	ECBOT-SOYBEAN OIL	Daily Factor: Commodities
$G_{2,D,t}$	A Rated minus Y10 Tbond	Daily Factor: Corporate Risk
$G_{3,D,t}$	6-month US T-bill	Daily Factor: Treasuries
$G_{4,D,t}$	6-month treasury bill minus 3M-Tbills	Daily Factor: Treasuries
$G_{5,D,t}$	CBT-MINI DOW JONES	Daily Factor: Equities
$G_{6,D,t}$	Corn	Daily Factor: Commodities
$G_{7,Dt}$	APFNF-AAF	Daily Factor: Corporate Risk
$G_{8,D,t}$	US nominal dollar broad index	Daily Factor: FX

Note: This table reports the series with largest marginal R^2 for the factor specified in the first column. The marginal R^2 is computed from regressions of each of the individual series onto the factor, one at a time, for the time period that the factor shows up as relevant for the median bias.

Predictor Variables

Let $\mathbf{Z}_{jt} \equiv \left(y_{j,t}, \hat{\mathbf{G}}'_{jt}, \mathbf{W}'_{jt}\right)'$ be a $r = 1 + r_G + r_W$ vector which collects $y_{j,t}$, the r_G estimated factors, and the r_W additional predictors, and define $\mathcal{Z}_{jt} \equiv \left(y_{j,t}, ..., y_{j,t-p_y}, \hat{\mathbf{G}}'_{jt}, ..., \hat{\mathbf{G}}'_{jt-p_G}, \mathbf{W}'_{jt}, ..., \mathbf{W}'_{jt-p_W}\right)'$, where p_y, p_G, p_W are lags of $y_{j,t}$, $\hat{\mathbf{G}}'_{jt}$, $\mathbf{W}'_{jt}$, respectively. We consider the following machine forecasting regression

$$y_{j,t+h} = \alpha_j + \beta_{j\mathbb{F}} \mathbb{F}_t^{(i)} \left[y_{j,t+h}\right] + \underbrace{\mathbf{B}_{j\mathcal{Z}}}_{1\times q} \mathcal{Z}_{jt} + \epsilon_{jt+h}$$

where $q = r + p_y + p_G + p_W$. Let superscript (i) refer to the ith forecaster, where i denotes either the mean "*mean*" or an ith percentile value of the forecast distribution, i.e., "65" is the 65th percentile. The predictors below are listed as elements of $y_{j,t}$, $\hat{\mathbf{G}}'_{jt}$, or $\mathbf{W}'_{jt}$ for different surveys and variables.

SPF Inflation For y_j equal to inflation the forecasting model considers the following variables.

In $\mathbf{W}'_{jt}$:

1. $\mathbb{F}_{jt-k}^{(i)} \left[y_{jt+h-k}\right]$, where $k = 1, \ldots, 2$
2. $\mathbb{F}_{jt-1}^{(s\neq i)} \left[y_{jt+h-1}\right]$, where $s = mean$, 50, 25, 75 for all $s \neq i$
3. $\text{var}_N \left(\mathbb{F}_{t-1}^{(\cdot)} \left[y_{jt+h-1}\right]\right)$, where $\text{var}_N (\cdot)$ denotes the cross-sectional variance of lagged survey forecasts
4. $\text{skew}_N \left(\mathbb{F}_{t-1}^{(\cdot)} \left[y_{jt+h-1}\right]\right)$, where $\text{skew}_N (\cdot)$ denotes the cross-sectional skewness of lagged survey forecasts
5. Trend inflation measured as $\overline{\pi}_{t-1} = \begin{cases} \rho\overline{\pi}_{t-2} + (1-\rho)\pi_{t-1}, \ \ \rho = 0.95 & \text{if } t < \text{1991:Q4} \\ \text{CPI10}_{t-1} & \text{if } t \geq \text{1991:Q4} \end{cases}$ Trend inflation is intended to capture long-run trends. When long-run forecasts of inflation are not available, as is the case pre-1991:Q4, we us a moving average of past inflation.
6. $\widetilde{GDP}_{t-1}$ = detrended gross domestic product, defined as the residual from a regression of GDP_{t-1} on a constant and the four most recent values of GDP as of date $t-8$. See Hamilton (2018).
7. $\widetilde{EMP}_{t-1}$ = detrended employment, defined as the residual from a regression of EMP_{t-1} on a constant and the four most recent values of EMP as of date $t-8$. See Hamilton (2018).
8. $\mathbb{N}_t^{(i)} \left[\pi_{t,t-h}\right]$ = Nowcast as of time t of the ith percentile of inflation over the period $t-h$ to t.

Lags of the dependent variable:

1. $y_{t-1,t-h-1}$ one quarter lagged annual inflation.

The factors in $\hat{\mathbf{G}}'_{jt}$ include factors formed from three large datasets separately:

1. $\mathbf{G}_{M,t-k}$, for $k = 0, 1$ are factors formed from a real time macro dataset $\mathcal{D}^M$ with 92 real time macro series; includes both monthly and quarterly series, with monthly series converted to quarterly according to the method described in the data appendix.

2. $\mathbf{G}_{F,t-k}$, for $k = 0, 1$ are factors formed from a financial data set $\mathcal{D}^F$ with 147 monthly financial series.

3. $\mathbf{G}^Q_{D,t}$, are quarterly factors formed from a daily financial dataset $\mathcal{D}^D$ of 87 daily financial indicators. The raw daily series are first converted to daily factors $\mathbf{G}_{D,t}(\mathbf{w})$ and the daily factors are aggregated up to quarterly observations $\mathbf{G}^Q_{D,t}(\mathbf{w})$ using a weighted average of daily factors, with the weights $\mathbf{w}$ dependent on two free parameters that are chosen to minimize the sum of squared residuals in a regression of $y_{j,t+h}$ on $\mathbf{G}_{D,t}(\mathbf{w})$.

The 92 macro series in $\mathcal{D}^M$ are selected to represent broad categories of macroeconomic time series. The majority of these are real activity measures: real output and income, employment and hours, consumer spending, housing starts, orders and unfilled orders, compensation and labor costs, and capacity utilization measures. The dataset also includes commodity and price indexes and a handful of bond and stock market indexes, and foreign exchange measures. The financial dataset $\mathcal{D}^f$ is an updated monthly version of the of 147 variables comprised solely of financial market time series used in Ludvigson and Ng (2007). These data include valuation ratios such as the dividend-price ratio and earnings-price ratio, growth rates of aggregate dividends and prices, default and term spreads, yields on corporate bonds of different ratings grades, yields on Treasuries and yield spreads, and a broad cross-section of industry, size, book-market, and momentum portfolio equity returns.[14] The 87 daily financial indicators in $\mathcal{D}^D$ include daily time series on commodities spot prices and futures prices, aggregate stock market indexes, volatility indexes, credit spreads and yield spreads, and exchange rates.

SPF GDP Growth For y_j equal to GDP growth the forecasting model considers the following variables.

In $\mathbf{W}'_{jt}$

1. $\mathbb{F}^{(i)}_{jt-k}[y_{jt+h-k}]$, where $k = 1, 2$

2. $\mathbb{F}^{(s\neq i)}_{jt-1}[y_{jt+h-1}]$, where $s = mean$, 50, 25, 75 for all $s \neq i$

3. $\mathrm{var}_N\left(\mathbb{F}^{(\cdot)}_{t-1}[y_{jt+h-1}]\right)$, where $\mathrm{var}_N(\cdot)$ denotes the cross-sectional variance of forecasts

4. $\mathrm{skew}_N\left(\mathbb{F}^{(\cdot)}_{t-1}[y_{jt+h-1}]\right)$, where $\mathrm{skew}_N(\cdot)$ denotes the cross-sectional skewness of forecasts

[14] A detailed description of the series is given in the Data Appendix of the online supplementary file at www.sydneyludvigson.com/s/ucc_data_appendix.pdf

5. $\widetilde{GDP}_{t-1}$ = detrended gross domestic product, defined as the residual from a regression of GDP_{t-1} on a constant and the four most recent values of GDP as of date $t-8$. See Hamilton (2018).

6. $\widetilde{EMP}_{t-1}$ = detrended employment, defined as the residual from a regression of EMP_{t-1} on a constant and the four most recent values of EMP as of date $t-8$. See Hamilton (2018).

7. $\mathbb{N}_t^{(i)}[y_{t,t-h}]$ = Nowcast as of time t of the ith percentile of GDP growth over the period $t-h$ to t.

8. VXO_t, defined as CBOE S&P 100 volatility index. We also include its squared and cubic terms, VXO_t^2, and VXO_t^3.

Lags of the dependent variable:

1. $y_{j,t-1,t-h-1}, y_{j,t-2,t-h-2}$ one and two quarter lagged annual GDP growth.

The factors in $\hat{\mathbf{G}}'_{jt}$ include factors formed from three large datasets separately:

1. $\mathbf{G}_{M,t-k}$, for $k = 0, 1$ are factors formed from a real time macro dataset $\mathcal{D}^M$ with 92 real time macro series; includes both monthly and quarterly series, with monthly series converted to quarterly according to the method described in the data appendix.

2. $\mathbf{G}_{F,t-k}$, for $k = 0, 1$ are factors formed from a financial data set $\mathcal{D}^F$ with 147 monthly financial series.

3. $\mathbf{G}_{D,t}^Q$, are quarterly factors formed from a daily financial dataset $\mathcal{D}^D$ of 87 daily financial indicators. The raw daily series are first converted to daily factors $\mathbf{G}_{D,t}(\mathbf{w})$ and the daily factors are aggregated up to quarterly observations $\mathbf{G}_{D,t}^Q(\mathbf{w})$ using a weighted average of daily factors, with the weights $\mathbf{w}$ dependent on two free parameters that are chosen to minimize the sum of squared residuals in a regression of $y_{j,t+h}$ on $\mathbf{G}_{D,t}(\mathbf{w})$.

The 92 macro series in $\mathcal{D}^M$ are selected to represent broad categories of macroeconomic time series. The majority of these are real activity measures: real output and income, employment and hours, consumer spending, housing starts, orders and unfilled orders, compensation and labor costs, and capacity utilization measures. The dataset also includes commodity and price indexes and a handful of bond and stock market indexes, and foreign exchange measures. The financial dataset $\mathcal{D}^f$ is an updated monthly version of the of 147 variables comprised solely of financial market time series used in Ludvigson and Ng (2007). These data include valuation ratios such as the dividend-price ratio and earnings-price ratio, growth rates of aggregate dividends and prices, default and term spreads, yields on corporate bonds of different ratings grades, yields on Treasuries and yield spreads, and a broad cross-section of industry, size, book-market, and momentum portfolio equity returns.[15] The 87 daily financial indicators in $\mathcal{D}^D$ include daily time series on commodities spot prices and futures prices, aggregate stock market indexes, volatility indexes, credit spreads and yield spreads, and exchange rates.

[15] A detailed description of the series is given in the Data Appendix of the online supplementary file at www.sydneyludvigson.com/s/ucc_data_appendix.pdf

SOC Inflation For consistency, the predictors for the SOC inflation forecasts are constructed similarly to those of the SPF inflation forecasts. Again, consider the following forecast regression,

$$y_{j,t+h} = \alpha_j + \beta_{j\mathbb{F}} \mathbb{F}^{MS,(i)}_{j,t} [y_{j,t+h}] + \underbrace{\mathbf{B}_{j\mathcal{Z}}}_{1xq} \mathcal{Z}_{jt} + \epsilon_{jt+h},$$

where the variables are defined as above, and i is either the mean "$mean$" or an ith percentile value of the forecast distribution. We denote forecasts from the SPF using $\mathbb{F}^{SPF,(i)}_{js} [\cdot]$ and from the Michigan Survey using $\mathbb{F}^{MS,(i)}_{js} [\cdot]$.

In $\mathbf{W}'_{jt}$:

1. $\mathbb{F}^{SPF,(\mu)}_{jt-1} [y_{jt+h-1}]$, the mean SPF forecast for CPI.
2. $\mathbb{F}^{SPF,(50)}_{jt-1} [y_{jt+h-1}]$, the $50th$ percentile SPF forecast for CPI.
3. $\mathbb{F}^{SPF,(25)}_{jt-1} [y_{jt+h-1}]$, the $25th$ percentile SPF forecast for CPI.
4. $\mathbb{F}^{SPF,(75)}_{jt-1} [y_{jt+h-1}]$, the $75th$ percentile SPF forecast for CPI.
5. $\text{var}_N \left(\mathbb{F}^{SPF,(\cdot)}_{t-1} [y_{jt+h-1}] \right)$, the cross-sectional variance of SPF forecasts of CPI.
6. $\text{skew}_N \left(\mathbb{F}^{SPF,(\cdot)}_{t-1} [y_{jt+h-1}] \right)$, the cross-sectional skewness of SPF forecasts of CPI.
7. Trend inflation measured as $\overline{\pi}_{t-1} = \begin{cases} \rho \overline{\pi}_{t-2} + (1-\rho)\pi_{t-1}, \;\; \rho = 0.95 & \text{if } t < \text{1991:Q4} \\ \text{CPI10}_{t-1} & \text{if } t \geq \text{1991:Q4} \end{cases}$ Trend inflation is intended to capture long-run trends. When long-run forecasts of inflation are not available, as is the case pre-1991:Q4, we us a moving average of past inflation.
8. $\widetilde{GDP}_{t-1}$ = detrended gross domestic product, defined as the residual from a regression of GDP_{t-1} on a constant and the four most recent values of GDP as of date $t-8$. See Hamilton (2018).
9. $\widetilde{EMP}_{t-1}$ = detrended employment, defined as the residual from a regression of EMP_{t-1} on a constant and the four most recent values of EMP as of date $t-8$. See Hamilton (2018).

Lags of dependent variables:

1. $y_{t-1,t-h-1}$ one quarter lagged annual CPI inflation.

The factors in $\hat{\mathbf{G}}'_{jt}$ include factors formed from three large datasets separately:

1. $\mathbf{G}_{M,t-k}$, for $k = 0, 1$ are factors formed from a real time macro dataset $\mathcal{D}^M$ with 92 real time macro series; includes both monthly and quarterly series, with monthly series converted to quarterly according to the method described in the data appendix.

2. $\mathbf{G}_{F,t-k}$, for $k = 0, 1$ are factors formed from a financial data set $\mathcal{D}^F$ with 147 monthly financial series.

3. $\mathbf{G}^Q_{D,t}$, are quarterly factors formed from a daily financial dataset $\mathcal{D}^D$ of 87 daily financial indicators. The raw daily series are first converted to daily factors $\mathbf{G}_{D,t}(\mathbf{w})$ and the daily factors are aggregated up to quarterly observations $\mathbf{G}^Q_{D,t}(\mathbf{w})$ using a weighted average of daily factors, with the weights $\mathbf{w}$ dependent on two free parameters that are chosen to minimize the sum of squared residuals in a regression of $y_{j,t+h}$ on $\mathbf{G}_{D,t}(\mathbf{w})$.

The 92 macro series in $\mathcal{D}^M$ are selected to represent broad categories of macroeconomic time series. The majority of these are real activity measures: real output and income, employment and hours, consumer spending, housing starts, orders and unfilled orders, compensation and labor costs, and capacity utilization measures. The dataset also includes commodity and price indexes and a handful of bond and stock market indexes, and foreign exchange measures. The financial dataset $\mathcal{D}^f$ is an updated monthly version of the of 147 variables comprised solely of financial market time series used in Ludvigson and Ng (2007). These data include valuation ratios such as the dividend-price ratio and earnings-price ratio, growth rates of aggregate dividends and prices, default and term spreads, yields on corporate bonds of different ratings grades, yields on Treasuries and yield spreads, and a broad cross-section of industry, size, book-market, and momentum portfolio equity returns.[16] The 87 daily financial indicators in $\mathcal{D}^D$ include daily time series on commodities spot prices and futures prices, aggregate stock market indexes, volatility indexes, credit spreads and yield spreads, and exchange rates.

SOC GDP Growth For y_j equal to GDP growth the forecasting model considers the following variables

In $\mathbf{W}'_{jt}$:

1. $\mathbb{F}^{SPF,(\mu)}_{jt-1}[y_{jt+h-1}]$, the mean SPF forecast for GDP growth.

2. $\mathbb{F}^{SPF,(50)}_{jt-1}[y_{jt+h-1}]$, the $50th$ percentile SPF forecast for GDP growth.

3. $\mathbb{F}^{SPF,(25)}_{jt-1}[y_{jt+h-1}]$, the $25th$ percentile SPF forecast for GDP growth.

4. $\mathbb{F}^{SPF,(75)}_{jt-1}[y_{jt+h-1}]$, the $75th$ percentile SPF forecast for GDP growth.

5. $\text{var}_N\left(\mathbb{F}^{SPF,(\cdot)}_{t-1}[y_{jt+h-1}]\right)$, the cross-sectional variance of SPF forecasts for GDP growth.

6. $\text{skew}_N\left(\mathbb{F}^{SPF,(\cdot)}_{t-1}[y_{jt+h-1}]\right)$, the cross-sectional skewness of SPF forecasts for GDP growth.

7. $\widetilde{GDP}_{t-1}$ = detrended gross domestic product, defined as the residual from a regression of GDP_{t-1} on a constant and the four most recent values of GDP as of date $t-8$. See Hamilton (2018).

[16]A detailed description of the series is given in the Data Appendix of the online supplementary file at www.sydneyludvigson.com/s/ucc_data_appendix.pdf

8. $\widetilde{EMP}_{t-1}$ = detrended employment, defined as the residual from a regression of EMP_{t-1} on a constant and the four most recent values of EMP as of date $t-8$. See Hamilton (2018).

9. VXO_t, defined as CBOE S&P 100 volatility index. We also include its squared and cubic terms, VXO_t^2, and VXO_t^3.

Lags of dependent variables:

1. $y_{j,t-1,t-h-1}, y_{j,t-2,t-h-2}$ one and two quarter lagged annual GDP growth.

The factors in $\hat{\mathbf{G}}'_{jt}$ include factors formed from three large datasets separately:

1. $\mathbf{G}_{M,t-k}$, for $k = 0, 1$ are factors formed from a real time macro dataset $\mathcal{D}^M$ with 92 real time macro series; includes both monthly and quarterly series, with monthly series converted to quarterly according to the method described in the data appendix.

2. $\mathbf{G}_{F,t-k}$, for $k = 0, 1$ are factors formed from a financial data set $\mathcal{D}^F$ with 147 monthly financial series.

3. $\mathbf{G}^Q_{D,t}$, are quarterly factors formed from a daily financial dataset $\mathcal{D}^D$ of 87 daily financial indicators. The raw daily series are first converted to daily factors $\mathbf{G}_{D,t}(\mathbf{w})$ and the daily factors are aggregated up to quarterly observations $\mathbf{G}^Q_{D,t}(\mathbf{w})$ using a weighted average of daily factors, with the weights $\mathbf{w}$ dependent on two free parameters that are chosen to minimize the sum of squared residuals in a regression of $y_{j,t+h}$ on $\mathbf{G}_{D,t}(\mathbf{w})$.

The 92 macro series in $\mathcal{D}^M$ are selected to represent broad categories of macroeconomic time series. The majority of these are real activity measures: real output and income, employment and hours, consumer spending, housing starts, orders and unfilled orders, compensation and labor costs, and capacity utilization measures. The dataset also includes commodity and price indexes and a handful of bond and stock market indexes, and foreign exchange measures. The financial dataset $\mathcal{D}^f$ is an updated monthly version of the of 147 variables comprised solely of financial market time series used in Ludvigson and Ng (2007). These data include valuation ratios such as the dividend-price ratio and earnings-price ratio, growth rates of aggregate dividends and prices, default and term spreads, yields on corporate bonds of different ratings grades, yields on Treasuries and yield spreads, and a broad cross-section of industry, size, book-market, and momentum portfolio equity returns.[17] The 87 daily financial indicators in $\mathcal{D}^D$ include daily time series on commodities spot prices and futures prices, aggregate stock market indexes, volatility indexes, credit spreads and yield spreads, and exchange rates.

Blue Chip Inflation For consistency, the predictors for the BC inflation (PGDP inflation and CPI inflation) forecasts are constructed analogously to those of the SPF inflation forecasts. The only differences are that for own-survey forecasting variables (including nowcasts), e.g. $\mathbb{F}_t^{(i)}[y_{jt+h}]$ in $\mathbf{W}'_{jt}$, we now use survey forecasts from Blue Chip, instead of SPF.

[17]A detailed description of the series is given in the Data Appendix of the online supplementary file at www.sydneyludvigson.com/s/ucc_data_appendix.pdf

Blue Chip GDP Growth For y_j equal to GDP growth the forecasting model considers the same variables as in the SPF GDP growth forecasts with SPF forecasts replaced with Blue Chip Forecasts.

Coibion Gorodnichenko Regressions

To construct SPF forecasts of annual inflation, forecasters at time t are presumed to use an advance estimate of $t-1$ price level combined with their survey respondent forecast of that price level at $t+3$ to form a forecast of π_{t+3}.

$$\underbrace{\pi_{t+3} - \mathbb{F}_t^{(\mu)}\left[\pi_{t+3}\right]}_{\text{Forecast Error}} = \alpha + \beta\left(\underbrace{\mathbb{F}_t^{(\mu)}\left[\pi_{t+3}\right] - \mathbb{F}_{t-1}^{(\mu)}\left[\pi_{t+3}\right]}_{\text{Forecast Revision}}\right) + \epsilon_{t+3} \tag{A.14}$$

where the annual inflation at time $t+3$ is defined as,

$$\pi_{t+3} = 100 \times \left(\frac{P_t}{P_{t-1}} \times \frac{P_{t+1}}{P_t} \times \frac{P_{t+2}}{P_{t+1}} \times \frac{P_{t+3}}{P_{t+2}} - 1\right). \tag{A.15}$$

Following CG, regressions are run and forecast errors computed using forecasts of real-time inflation data available four quarters after the period being forecast.

The survey forecast is constructed as follows

$$\mathbb{F}_t\left[\pi_{t+3}\right] = 100 \times \left(\frac{P_t^{avg}}{P_{t-1}} \times \frac{P_{t+1}^{avg}}{P_t^{avg}} \times \frac{P_{t+2}^{avg}}{P_{t+1}^{avg}} \times \frac{P_{t+3}^{avg}}{P_{t+2}^{avg}} - 1\right),$$

where $P_{t+h}^{avg} = \frac{1}{N_{t+h}} \sum_{i=1}^{N_{t+h}} P_{t+h}^i$, for $h = 0, \ldots, 3$, i represents an individual forecaster, N_{t+h} is the number of forecasters at time time $t+h$, and P_{t-1} is the BEA's advance estimate at t for prices in $t-1$.

Forecast Error

The forecast error on the LHS of the regressions (A.14) is constructed in the following way:

$$\begin{aligned}\pi_{t+3,t} - \mathbb{F}_t^{(\mu)}\left[\pi_{t+3,t}\right] \equiv 100 \times \Bigg[& \left(\frac{\pi_{t,t-1} - \mathbb{F}_t^{(\mu)}\left[\pi_{t,t-1}\right]}{400} + 1\right) \\ & \times \left(\frac{\pi_{t+1,t} - \mathbb{F}_t^{(\mu)}\left[\pi_{t+1,t}\right]}{400} + 1\right) \\ & \times \left(\frac{\pi_{t+2,t+1} - \mathbb{F}_t^{(\mu)}\left[\pi_{t+2,t+1}\right]}{400} + 1\right) \\ & \times \left(\frac{\pi_{t+3,t+2} - \mathbb{F}_t^{(\mu)}\left[\pi_{t+3,t+2}\right]}{400} + 1\right) - 1\Bigg]\end{aligned} \tag{A.16}$$

In brackets is the product of quarterly forecast errors from the nowcast to $h = 3$ quarters ahead.

In-sample analysis

Table A.6 presents the replication for CG, as well as results from extending the sample size to 2018:Q2. Panel A replicates the numbers from columns (1) and (2) of Table 1 Panel B of CG. Panel B presents the results for the extended sample.

Table A.6: CG In-Sample Regressions of Forecast Errors on Forecast Revisions

Regression: $\pi_{t+3,t} - \mathbb{F}_t[\pi_{t+3,t}] = \alpha + \beta(\mathbb{F}_t[\pi_{t+3,t}] - \mathbb{F}_{t-1}[\pi_{t+3,t}]) + \delta\pi_{t+2,t-1} + \epsilon_t$				
	(1)	(2)	(3)	(4)
	Panel A: Sample: 1969:Q1 - 2014:Q4		**Panel B**: Sample: 1969:Q1 - 2018:Q2	
Constant	0.001	-0.077	-0.022	-0.116
t-stat	(0.005)	(-0.442)	(-0.167)	(-0.758)
$\mathbb{F}_t[\pi_{t+3,t}] - \mathbb{F}_{t-1}[\pi_{t+3,t}]$	1.194**	1.141**	1.186**	1.116**
t-stat	(2.496)	(2.560)	(2.478)	(2.532)
$\pi_{t+2,t-1}$		0.021		0.027
t-stat		(0.435)		(0.574)
$\bar{R}^2$	0.195	0.197	0.193	0.195

Notes: The annual inflation is defined as $\pi_{t+3,t} = \frac{P_t}{P_{t-1}} \times \frac{P_{t+1}}{P_t} \times \frac{P_{t+2}}{P_{t+1}} \times \frac{P_{t+3}}{P_{t+2}}$, the covariate $\mathbb{F}_t[\pi_{t+3,t}]$ is the SPF of annual inflation with information in period t and $\mathbb{F}_{t-1}[\pi_{t+3,t}]$ is the SPF mean forecast of the same annual inflation but with information in $t-1$. Panel A presents the sample in Coibion and Gorodnichenko (2015) and Panel B updates the sample to 2018:Q2. Regressions are run and model evaluated using real-time data with observation on $\pi_{t+3,t}$ available 4 quarters after the advance estimate of it. Newey-West corrected (t-statistics) with lags = 4. Newey-West HAC: *sig. at 10%. **sig. at 5%. ***sig. at 1%.

Out-of-Sample Analysis

We seek to construct a series of real-time OOS forecasts using the model:

$$\pi_{t+3} - \mathbb{F}_t^{(\mu)}[\pi_{t+3}] = \alpha^{(\mu)} + \beta^{(\mu)}\left(\mathbb{F}_t^{(\mu)}[\pi_{t+3}] - \mathbb{F}_{t-1}^{(\mu)}[\pi_{t+3}]\right) + \epsilon_{t+3}$$

We estimate over an initial sample, forecast out one period, roll (or recurse) forward and repeat estimation and forecast. The regression estimation uses the latest vintage of inflation in real time and, following CG, computes forecast errors real-time data available four quarters after the period being forecast. The CG model forecast for π_{t+3}

$$\widehat{\pi}_{t+3}^{(\mu)} = \widehat{\alpha}_t^{(\mu)} + \left(1 + \widehat{\beta}_t^{(\mu)}\right)\mathbb{F}_t^{(\mu)}[\pi_{t+3}] - \widehat{\beta}_t^{(\mu)}\mathbb{F}_{t-1}^{(\mu)}[\pi_{t+3}]$$

For the rolling procedure, we try windows of sizes $w = 5$, 10, and 20 years. For the recursive procedure, we try initial window sizes of 5, 10, and 20 years as well.

The survey and model errors are

$$\text{survey error}_t = \mathbb{F}_t^{(\mu)}[\pi_{t+3}] - \pi_{t+3}$$
$$\text{CG model error}_t = \widehat{\pi}_{t+3}^{(\mu)} - \pi_{t+3}$$

We also compute rolling MSEs over different forecast samples of size P as

$$\mathrm{MSE}_{\mathbb{F}} = \frac{1}{P}\sum_{s=1}^{P}\left(\text{survey error}_{t+s}\right)^2$$

$$\mathrm{MSE}_{\mathrm{CG}} = \frac{1}{P}\sum_{s=1}^{P}\left(\text{CG model error}_{t+s}\right)^2$$

Table A.7: Mean Square Errors for the CG Model and SPF

Forecast model: $\widehat{\pi}^{(\mu)}_{t+3} = \widehat{\alpha}^{(\mu)}_t + \left(1+\widehat{\beta}^{(\mu)}_t\right) F^{(\mu)}_t\left[\pi_{t+3}\right] - \widehat{\beta}^{(\mu)}_t F^{(\mu)}_{t-1}\left[\pi_{t+3}\right]$		
	$\mathrm{MSE}_{CG}/\mathrm{MSE}_F$	
Method	Quarterly Compound	Continuous Compound
Rolling 5 years	1.38	1.38
Rolling 10 years	1.29	1.29
Rolling 20 years	1.31	1.30
Recursive 5 years	1.69	1.68
Recursive 10 years	1.60	1.59
Recursive 20 years	1.33	1.30

Notes: The table reports the ratio of MSEs of the CG model forecast over the survey forecast. The regression estimation uses the lastest vintage of inflation in real time and, following CG, computes forecast errors real-time data available four quarters after the period being forecast. The sample spans the period 1969:Q1 - 2018:Q2.