DISCUSSION PAPER SERIES

DP15003
(v. 4)

Belief Distortions and Macroeconomic Fluctuations

Francesco Bianchi, Sydney Ludvigson and Sai Ma

MONETARY ECONOMICS AND FLUCTUATIONS

ISSN 0265-8003

Belief Distortions and Macroeconomic Fluctuations

Francesco Bianchi, Sydney Ludvigson and Sai Ma

Discussion Paper DP15003
First Published 06 July 2020
This Revision 27 July 2021

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

Belief Distortions and Macroeconomic Fluctuations

Abstract

This paper combines a data rich environment with a machine learning algorithm to provide new estimates of time-varying systematic expectational errors ("belief distortions") embedded in survey responses. We find that distortions are large even for professional forecasters, with all respondent-types over-weighting their own beliefs relative to publicly available information. Forecasts of inflation and GDP growth oscillate between optimism and pessimism by large margins, with biases in expectations evolving dynamically in response to cyclical shocks. The results suggest that artificial intelligence algorithms can be productively deployed to correct errors in human judgement and improve predictive accuracy.

JEL Classification: E7, E27, E32, E17, G4

Keywords: beliefs, Biases, Machine Learning, Expectations

Francesco Bianchi - francesco.bianchi@jhu.edu
Duke University and CEPR

Sydney Ludvigson - sydney.ludvigson@nyu.edu
NYU and CEPR

Sai Ma - sai.ma@frb.gov
Board of Governors of the FRS

Belief Distortions and Macroeconomic Fluctuations

Francesco Bianchi*
Johns Hopkins, Duke
CEPR, and NBER

Sydney C. Ludvigson†
NYU, CEPR, and NBER

Sai Ma‡
Federal Reserve Board

First draft: April 2017
This draft: July 27, 2021

Abstract

This paper combines a data rich environment with a machine learning algorithm to provide new estimates of time-varying systematic expectational errors ("belief distortions") embedded in survey responses. We find that distortions are large even for professional forecasters, with all respondent-types over-weighting their own beliefs relative to publicly available information. Forecasts of inflation and GDP growth oscillate between optimism and pessimism by large margins, with biases in expectations evolving dynamically in response to cyclical shocks. The results suggest that artificial intelligence algorithms can be productively deployed to correct errors in human judgement and improve predictive accuracy.

JEL: E7, E27, E32, E17, G4.
Keywords: Beliefs, biases, machine learning, expectations

*Department of Economics, Duke, 213 Social Sciences building, Box 90097, Durham, NC 27708. (francesco.bianchi@duke.edu)

†Department of Economics, NYU, 19 W.4th St, 6th Floor, New York, NY 10012. (sydney.ludvigson@nyu.edu)

‡Federal Reserve Board of Governors, C Ave & 20th Street NW, Washington, DC 20551. (sai.ma@frb.gov)
We thank Marios Angeletos and Fabrice Collard for providing data on their estimated cyclical shocks, and Michael Boutros, Josue Cox, Justin Shugarman, and Yueteng Zhu for excellent research assistance. We are grateful to Marios Angeletos, Rudi Bachmann, Fabrice Collard, Andrew Foerster, Xavier Gabaix, David Hershleifer, Cosmin Ilut, Anil Kashyp, Laura Veldkamp, and to seminar participants at the Bank of Israel, Chicago Booth, Duke, the Federal Reserve Board, Richmond Fed, UC Berkeley, 2021 AEA Meeting, the July 2020 NBER Behavioral Macro workshop, the New Approaches for Modeling Expectations in Economics Conference (London), 2019, the III Conference on Applied Macro-Finance (Melbourne), 2019, 2020 Fed System Econometrics Meeting, King's Business School, and the 2020 Stanford Institute for Theoretical Economics Workshop on Asset Pricing, Macro Finance, and Computation for many helpful comments. The views expressed are those of the authors and do not necessarily reflect those of the Federal Reserve Board or the Federal Reserve System.

1 Introduction

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

To formalize our notion of "belief distortion," let us define it in general terms as *an ex ante 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 macroeconomic literatures where distorted beliefs play a role, economic agents make systematic expectational errors due to a wide variety of reasons. These include the presence of information frictions driven by rational or behavioral inattention, the use of simple extrapolative rules, the intentional adoption of conservatively pessimistic beliefs, the over-reaction to incoming news, or the presence of skewed priors, among others.

In this study we ask: How distorted are observed beliefs about the macroeconomy? Do any such distortions vary with the business cycle? Answers to these questions are inextricably tied to the measurement of belief distortions.

A fundamental challenge in this regard is that no objective measure of such distortions exists. So far, empirical work has largely proceeded by investigating whether forecast errors made by survey respondents deviate from the standard of full information and rational expectations. Yet a review of the literature discussed below finds little agreement on how such a theoretical standard should be measured. Existing 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 proposes 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. A general premise of our approach is that big data algorithms can be productively employed to reveal subjective biases in human judgements.

Once we have a method for uncovering those biases, artificial intelligence algorithms can be deployed to "correct" those errors and improve predictive accuracy.

Returning to our definition of belief distortions above, it is clear that measuring the errors in human judgements requires four 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. For this, we use tools for data rich environments along with machine learning 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. Third, we must account for other bona fide features of real time decision making, such as the out-of-sample nature of forward-looking judgements. Failure to properly account for 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. Conversely, using information that may have been *un*available to survey respondents to compute a standard of non-distorted beliefs could be equally erroneous. To address these issues, we develop a dynamic machine learning algorithm explicitly designed to combat overfitting in order to detect demonstrable, *ex ante* expectational errors in real-time. The fourth 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 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 relation of any belief distortions to cyclical fluctuations.

With these ingredients in hand, we ask whether cross-sections of survey respondents with different beliefs systematically mis-weight pertinent economic information. If the machine detects a sustained pattern of demonstrable, *ex ante* errors in survey respondents' forecasts, the magnitude of these distortions should be evident from the relative (machine versus respondent-type) out-of-sample forecast errors once averaged over a sample sufficiently long so as to eliminate differences in *ex post* predictive outcomes attributable to random 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 at least two 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 even households. Machine algorithms are advantageous in this regard because they are explicitly designed to cope with large amounts of information. This is important because 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. Second, a machine algorithm can be coded to systematically adapt to new information as it becomes available and to make out-of-sample forecasts on this basis. Thus the approach does not run the risk of spuriously indicating that respondent performance is suboptimal merely because of the existence of structural breaks and/or the arrival of new information that even an efficient information processing algorithm could have learned about only slowly over time. More generally, we argue that the machine-based methods offer hope for improving prediction and estimation in a range of settings that rely on human surveys as a major empirical input.

Inherent in our machine-based approach is the idea that minding key features of real world expectation formation is essential when establishing a benchmark against which belief distortions are measured. Whether doing so matters in practice, however, is an empirical question. On this question, we can report at least three ways in which our results differ from some in the extant literature. First, in contrast to well known results from in-sample regressions, we find little evidence that lagged *ex ante* revisions in survey forecasts have predictive power for average survey forecast errors. Second, information found elsewhere to be consequential for out-of-sample prediction in a low-dimensional setting is often found to be unimportant in our high-dimensional, data rich setting. Third, measures of belief distortions created by comparing *ex ante* survey expectations with theoretical benchmarks that rely on *ex post* historical outcome data overstate the magnitude of distortion.

Our main economic findings may be summarized as follows. First, across a range of surveys, variables, and respondent-types with heterogeneous beliefs, the machine model produces lower mean squared forecast error over long external evaluation samples, sometimes by large margins. The magnitude of improvement is especially large in the last five years of the sample, from 2013:Q2-2018:Q2. A key finding is that survey respondents of all types often place too much weight on the marginal information embedded in their own forecasts and too little weight on objective, publicly available economic information, a finding suggestive of an overreliance by survey respondents on the private or judgemental component of their forecasts. Below we present a simple framework of public and private signals that facilitates this interpretation.

Second, survey expectations of inflation for the median respondent of all surveys are biased upward on average, a direction we shall refer to as "pessimistic." By contrast, survey expectations of economic growth by professional forecasters and corporate executives are "optimistic" on average–i.e., biased upward, while they are very slightly pessimistic for households. These biases are found to be largest at the end of our sample, from 2013:Q2-2018:Q2, when the median forecast of economic growth from the Survey of Professional Forecasters (SPF) was biased upward by an amount equal to 34% of actual GDP growth over this period, resulting in forecasts that were 30% less accurate on average than the machine specification. The median SPF forecaster also persistently over-estimated inflation during this time period, resulting in forecasts that were 42% less accurate than the machine specification.

Third, although our machine learning algorithm indicates that sparse specifications are often optimal, this is not the case in every period. Moreover, even with sparse specifications, the precise information utilized changes from period to period. These results underscore the importance of using a dynamic, large-scale information processing algorithm to reduce errors in human judgement, even if much of the information the algorithm considers is associated with a coefficient that is shrunk all the way to zero most of the time.

Fourth, although the machine is able to detect patterns in the data that notably improve predictive accuracy over human forecasts, these improvements are best described as incremental and certainly produce smaller estimates of belief distortion than suggested by some previous empirical studies, as discussed below. In contrast to these studies, however, our benchmark of non-distorted beliefs is formed by (i) relying exclusively on information that we can verify could have been known to survey respondents on or before the survey response deadline, (ii) requiring the machine to choose dynamically evolving, best-fitting empirical specifications *ex ante* rather than with hindsight, and (iii) employing genuine out-of-sample prediction in the external validation step. The strict adherence to these principles means that we find a smaller magnitude of belief distortion in the real-time predictions of professional forecasters, and that there are times when the machine invariably makes large *ex post* mistakes. A notable example is the Great Recession, which the machine failed to recognize in real time, resulting in big forecast errors similar in magnitude to those made by professional forecasters during this episode. We argue that such episodes underscore the role of largely unforeseen events in generating occasionally large prediction error, not all of which can be attributed to a systematic bias in expectations.

The rest of this paper is organized as follows. Section 2 reviews the related literature. Section 3 describes our econometric and machine learning framework. Section 4 describes results pertaining to our estimates of belief distortions and relates our estimates with those from approaches used in some well-known prior empirical studies. Section 5 contains results on how belief distortions change over the business cycle. Section 6 concludes. A large amount of additional material on our data construction, estimation procedures, and additional robustness checks have been placed in an Appendix for online publication.

2 Related Literature

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.

In these theoretical literatures, economic agents make systematic expectational errors for a variety of reasons. These reasons include the presence of information frictions that lead agents to act in a "boundedly rational" manner because they are incapable of attending to all the available information at a given moment (e.g., Mankiw and Reis (2002); Woodford (2002); Sims (2003); Reis (2006a, 2006b); Gabaix (2014)). Alternatively agents may be inattentive for broader behavioral reasons (e.g., Gabaix (2020)). A key implication of these theories, explored in well known work by Coibion and Gorodnichenko (2015), is that individuals *under*-react to objective economic information. Our finding that belief distortions for professional forecasters are larger at the end rather than the beginning of our sample suggests that forms of bounded rationality attributable solely to limitations in information processing capacity are unlikely to fully explain our results. Similarly, our estimates are little changed if we allow the machine to observe every respondent-type's current forecast, suggesting that information frictions based on noisy "dispersed information" are also unlikely to be the most relevant source of systematic error we uncover.

Other theories postulate that individuals use simple extrapolative rules or over-weight "representative" events in reacting to incoming news (e.g., De Long, Shleifer, Summers, and Waldmann (1990); Barberis, Shleifer, and Vishny (1998); Barberis, Greenwood, Jin, and Shleifer (2015); Bordalo, Gennaioli, and Shleifer (2018); Gennaioli and Shleifer (2018); Bordalo, Gennaioli, Ma, and Shleifer (2018)). Related theories propose that individuals overweight their personal experiences (e.g., Malmendier and Nagel (2011, 2015)). A key implication of many of these theories is that individuals *over*-react to objective information.

A literature on "sentiments" postulates that communication frictions can cause aggregate expectations to exhibit statistical biases (e.g., Angeletos and La'O (2013); Angeletos, Collard, and Dellas (2018b); Milani (2011, 2017)). Other models feature "confidence shocks," or ambiguity averse agents who are deliberately pessimistic on average (e.g., Hansen and Sargent (2008); Epstein and Schneider (2010); Ilut and Schneider (2015); Bianchi, Ilut, and Schneider (2017); Ilut and Saijo (2020); Bhandari, Borovicka, and Ho (2019)), or agents with skewed priors (Afrouzi and Veldkamp (2019)). There remains a question of whether ambiguity aversion or skewed priors would be revealed in survey responses. If not, such models need some other mechanism to explain the systematic expectational errors documented here and elsewhere.

Finally a theoretical literature in economic psychology studies how basic properties of cognition can give rise to human biases in expectation formation (e.g., Woodford (2013); Khaw, Stevens, and Woodford (2017)).

Any of the above theories provide a mechanism through which a relatively unbiased and potentially more information-efficient machine operating in a data rich environment would provide forecasts that deviate from those made by humans and possibly be more accurate. The objective of this study is to provide new measures of such deviations and to investigate their relation to macroeconomic fluctuations.

On the empirical side, our work follows a growing body of literature that reports evidence of belief distortions and relates them to economic activity. These papers include those that 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 (2016); Bouchaud, Kruger, Landier, and Thesmar (2017)). Although these studies differ widely according to their empirical design, none take into account the data rich context in which survey respondents operate or the dynamic, out-of-sample nature of their forecasts, gaps our study is designed to fill.

These very differences lead our findings to diverge 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 that imply under-reaction to economic news. Using the methodology proposed in this paper, we find no evidence that they do. Instead, the coefficients on forecast revisions are shrunk to zero by the dynamic machine algorithm in favor of placing greater absolute weight on other pieces of information. Even if we use the same empirical specification used in Coibion and Gorodnichenko (2015), forecast revisions cease to be a useful predictors of forecast errors in a dynamic context when predictions are simply made out-of-sample rather than in-sample. Similarly, we ask whether survey respondents initially under-react to cyclical shocks but later over-react, as documented in Angeletos, Huo, and Sastry (2020). We confirm this general pattern but find that the magnitudes of under- and over-reaction are much smaller than those using the methodology of Angeletos et. al., in which the benchmark for measuring non-distorted beliefs is based on historical outcome data that would not have been known to survey respondents in real time.

The literature discussed so far has little to say about overconfidence, a term generally reserved in the behavioral economics literature to describe an agent who overestimates the precision of her private signal. Yet our finding that respondents of all types place too much weight on the marginal information embedded in their own forecasts is one of the most robust and quantitatively important contributors to bias that we uncover. Below we present a simple model of public and private signals to help interpret this finding. In this model, the machine will downweight the information contained in the survey response if the forecaster either over-estimates the precision of her private signal, as in traditional notions of overconfidence, and/or if she inefficiently combines the public information, thereby effectively under-estimating the precision of her public signal. Either way, the forecaster gives too much weight in relative terms to the private, judgemental component of her forecast. In this regard, our findings relate to extensive finance literature that provides theory and evidence of overconfidence and its role in explaining a range of stylized facts about stock return predictability and trading patterns. Ground breaking contributions include Odean (1998), Daniel, Hirshleifer, and Subrahmanyam (1998), Barber and Odean (2000) and Daniel, Hirshleifer, and Subrahmanyam (2001). Daniel and Hirshleifer (2015) provide an overview of this literature. More generally, our findings echo a large body of evidence in psychology showing that people–perhaps especially experts and professionals–give too much weight to their private judgements when making predictions (e.g., Kahneman, Sibony, and Sunstein (2021), Ch. 10). To the best of our knowledge, this paper is the first to find evidence suggesting that biases in professional macroeconomic predictions are partly attributable to a strong overreliance on the implicit judgemental component of their forecasts.

Our work also connects with a pre-existing econometric forecasting literature, which finds that survey forecasts of inflation are extremely difficult if not impossible to beat with statistical models in out-of-sample forecasting (e.g., 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)). Indeed, these studies conclude that the very best forecasts of inflation are the subjective ones provided by surveys. By contrast, our machine learning algorithm, with its focus on detecting demonstrable *ex ante* errors, performs better in out-of-sample forecasting than every percentile of all of the survey forecast distributions that we study.

Finally, we are aware of relatively little work that has used machine learning as a benchmark against which belief distortions are measured. 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 accounting for the interplay between a data rich environment and dynamic, out-of-sample forecasting generates findings about belief distortions that differ considerably from prior frameworks that side-step these aspects of real world decision making.

3 Econometric and Machine Learning Framework

This section describes our econometric and machine learning framework. This framework is applied to 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 and professional forecasters at 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.

3.1 Overview

Before getting into the details of our approach, we discuss two aspects of the general methodology.

First, because our investigation operates in a dynamic setting, the machine learning analysis we undertake requires a sufficiently long time series of observations, including those on survey responses. This is necessary because a pre-existing sample of time-series data is required to both estimate and train the machine prior to having it make its first out-of-sample forecast. Although the surveys themselves extend back far enough in time to accomplish this, the panel elements of the surveys are too limited to conduct the analysis on a respondent-level basis, since panelist participation is often brief and/or intermittent over time.[1] We therefore conduct the analysis for respondents of a particular *type* at time t, defined to be those in specific *percentiles* of the time t survey forecast distribution. A maintained assumption of our approach, explained in detail below, is that survey respondents know their own "type," so that they have a sense of where in the time t forecast distribution their belief is located. We argue that this assumption is a reasonable approximation to reality, at least for professional forecasters, many of whom routinely and continuously telegraph updates of their forecasts to clients and the press, while at the same time monitoring the evolving predictions of other forecasters at "rival" institutions. Such forecasters are therefore likely to have very good real-time information about their location in the professional forecast distribution.

To provide support for this claim, Figure 1 reports the median SPF forecast of four quarter ahead real GDP growth over time along with the median Bloomberg Consensus forecast of the same variable. The Bloomberg Consensus forecasts are published daily, and the observations shown in the figure are taken from the Bloomberg survey panel on the earliest day prior to the SPF survey deadline. The two series have a correlation of 93% over the sample, consistent with the idea that professional forecasters have access to reliable observable proxies for their time t location in the professional forecast distribution.

Furthermore, we argue that percentile-level analyses are likely to be a more plausible description of the empirical specifications actually employed by professional forecasters, given well documented caveats with the assignment of individual identification numbers to panelists who change their place of employment but remain in the survey. When panelists change places of employment, they often join entirely new forecasting teams with bespoke modeling practices and forecasting perspectives.[2] Panelists also go in and out of the surveys, with sometimes extended gaps in their participation. Under these circumstances the use of "types"–(i.e., where are we relative to consensus?)–is likely to be more natural in an econometric modeling context than specifications based on a sporadic history of an individual forecaster's own predictions.

We emphasize that our approach does not require, nor do we assume, that forecasters have a *time-invariant* type. As long as they know their contemporaneous type, the approach described

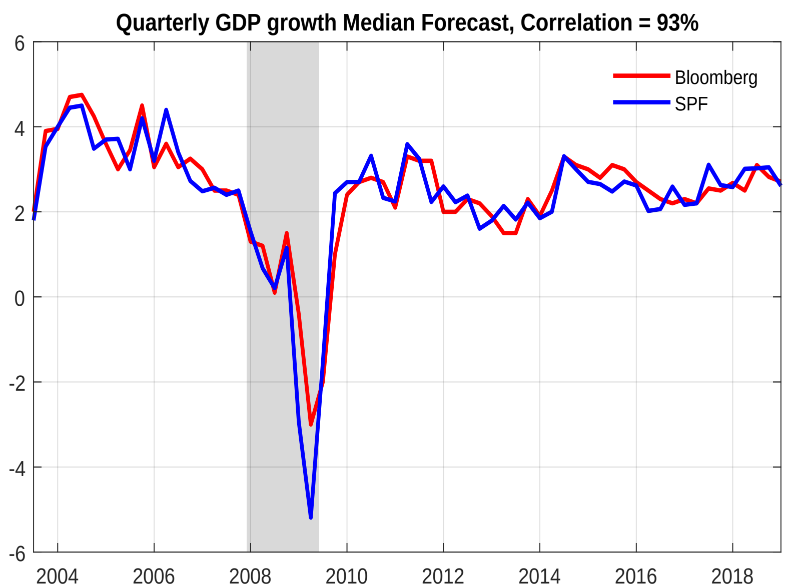
[1] The learning algorithm described below employs rolling estimation and training sample windows that could be as long as 34 quarters once combined, a span of data that must be available before the first out-of-sample machine forecast can be recorded. By contrast, the length of time that individual panelists remain in the survey samples is comparatively short. For example, for the Survey of Professional Forecasters survey on inflation expectations, the average forecaster remains in our sample just 18.5 quarters, with gaps in participation that would require filling in missing values.

[2] For example, see the memo "Caveats on Using the Individual Identification Numbers in the Survey of Professional Forecasters," posted at philedelphiafed.org.

below works even if respondents move around in the distribution. This follows because the one-period lagged values of every percentile's forecast are publicly available information, information that we find an efficient forecast would typically place non-zero weight on. Thus, a respondent can always ask how her time t type would have done historically by appending her current forecast onto the appropriate historical series.

With this in mind, we acknowledge that a core assumption of some economic theories is that individuals do not know their "rank" in the relevant cross-sectional distribution, in which case their "rank beliefs" can be important determinants of equilibria (e.g., Morris, Shin, and Yildiz (2016)). Below we assess the empirical relevance of such rank beliefs for our measured belief distortions, by comparing results using our baseline machine specifications with those from an alternative benchmark that uses the union of every type's forecast at time t.

Figure 1: Bloomberg and SPF Median Forecast



This figure reports the SPF median (blue) forecast and Bloomberg (red) median forecast of four quarter real GDP growth in percentage points, where the Bloomberg forecast is taken from the panel available on the closest day before the SPF survey deadline day. NBER recessions are shown with grey shaded bars. The sample spans 2003:Q1-2018:Q2.

A second aspect of the methodology pertains to our overall information processing approach. In order to identify possible distortions in beliefs, it is imperative that the benchmark model of belief formation use large and varied real time information sets, so that our measure of distortion does not miss pertinent information that could have been known to survey respondents, or pertain only to models with a small number of arbitrarily chosen information variables. To address this challenge, we take a two-pronged approach that combines diffusion index estimation with machine learning. Diffusion index forecasting, wherein a relatively small number of dynamic factors are estimated from hundreds of economic time-series, has become common in data-rich environments, following on a long line of prior studies showing that the approach improves prediction accuracy in a manner similar to model-averaging.[3] Diffusion indexes are also useful because some forms of nonlinearities are readily handled by including polynomial functions of estimated dynamic factors, or by forming additional factors from polynomials of the raw data. We use estimated factors as part of a dynamic machine learning algorithm of regularized estimation that chooses shrinkage and sparsity by optimally trading off the costs of downweighting information against the benefits of reduced parameter estimation error.[4] The diffusion index aspect of our methodology is standard, so we cover this step in the Online Appendix, focusing below on the dynamic machine learning framework.

3.2 Machine Efficient Benchmark

Let $y_{j,t+h}$ generically denote an 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 the ith *respondent-type*, where i denotes 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.

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. We assume that x_{it}^C has an approximate factor structure as detailed in the Online Appendix, where $\mathbf{G}_t^C$ is an $r_G \times 1$ vector of latent common factors ("diffusion indexes") with $\mathbf{\Lambda}_i^C$ a corresponding $r_C \times 1$ vector of latent factor loadings.

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

[3]An incomplete list of this literature includes Stock and Watson (1989);Stock and Watson (2006) Stock and Watson (1991); Stock and Watson (2002b); Stock and Watson (2002a);Stock and Watson (2006); Ludvigson and Ng (2007); Ludvigson and Ng (2009) and Ludvigson and Ng (2010).

[4]It is straightfoward to verify using hold out samples and/or artificial data that combining diffusion index estimation with machine learning is often better than doing either one in isolation, for two reasons. First, the optimal number of factors can still be large enough that there clear efficiency gains to using machine learning techniques for choosing shrinkage and sparsity even when all predictors are factors. Second, it is well known that the best approaches for choosing sparsity, such as those that use the L^1 "lasso" penalty, work poorly in the context of correlated regressors. Since our raw data are correlated, we have also verified that our elastic net estimator, which utilizes an L^1 penalty, works better when the data are first transformed into orthogonal diffusion indexes before estimation. This latter finding is consistent with results in Kozak, Nagel, and Santosh (2020).

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. Even with the use of factors, $\mathcal{Z}_{jt}$ can be of high dimension.

With these data in hand, consider the following machine learning empirical specification for forecasting $y_{j,t+h}$ given information at time t, to be benchmarked against the time t survey forecast of respondent-type i:

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

where $\alpha_j^{(i)}$, $\beta_{j\mathbb{F}}^{(i)}$, and $\mathbf{B}_{j\mathcal{Z}}^{(i)}$ are parameters to be estimated, and where $\mathbf{B}_{j\mathcal{Z}}^{(i)}$ is $K \times 1$, with $K = r + p_y + p_G \cdot r_G + p_W \cdot r_W$, the number of right-hand-side variables other than $\mathbb{F}_t^{(i)}$. Equation (1) is estimated using machine learning tools, as discussed below.

Estimation of (1) delivers a time t machine "belief" about $y_{j,t+h}$, namely the machine forecast, denoted $\mathbb{E}_t^{(i)} \left[y_{j,t+h} \right]$. We define the *machine efficient benchmark* as a set of parameter restrictions that would imply that 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 parameter restrictions:

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

Systematic expectational errors in the survey forecast are 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, $\mathbb{F}_t^{(i)} \left[y_{j,t+h} \right]$ (i.e., $\beta_{j\mathbb{F}}^{(i)} \neq 1$). Machine estimates $\widehat{\beta}_{j\mathbb{F}}^{(i)} \neq 1$ imply that the survey response $\mathbb{F}_t^{(i)} \left[y_{j,t+h} \right]$ could have been improved by giving it more or less weight relative to other objective economic information than the implicit weight of 1 given this response by the survey respondent. The machine can correct systematic errors in the human forecast by optimally re-weighting the marginal information contained in $\mathbb{F}_t^{(i)} \left[y_{j,t+h} \right]$ against the publicly available information contained in $\mathcal{Z}_{jt}$ that all survey respondents also had access to.

Three points about the machine efficient benchmark bear emphasis. First, it is a *type-specific* benchmark that adopts the perspective of a forecaster who is in the ith percentile of the survey forecast distribution in period t. The machine is given any information that the survey forecaster in the ith percentile could have observed at time t, including her own forecast $\mathbb{F}_t^{(i)} \left[y_{j,t+h} \right]$, as well as publicly available information contained in $\mathcal{Z}_{jt}$, where the latter includes lagged values of all other type's forecasts, since all surveys publish their results shortly after the response deadline. Once the machine specification controls for the information in $\mathcal{Z}_{jt}$, the marginal information in the survey response $\mathbb{F}_t^{(i)} \left[y_{j,t+h} \right]$ can only be that which is intangible to the machine, such as a private signal or judgement. Thus, allowing the machine to "see" the respondent type's time t forecast is crucial for capturing what role, if any, is played by the private or judgemental component of survey forecasts in generating belief distortions. In the next section we discuss a simple framework of private and public signals to facilitate this interpretation. We note, however, that even if we don't allow the machine to see the ith percentile's contemporaneous forecast and instead proxy for that observation using publicly available data from the Bloomberg Consensus forecast prior to t, our findings on forecaster bias are very similar.[5] But we argue that this case is far less interesting because it is silent on the possible role of private information or judgement in belief distortions.

Second, the machine is given only that information at time t that the survey respondent-type in the ith percentile could have observed at time t, and nothing more. This is important because superior machine forecasts formed with *ex post* information that we cannot be certain the survey respondent could have observed in real time might simply reflect the benefit of hindsight, rather than genuine systematic expectational error. For this reason, some popular techniques for forming benchmarks to measure forecaster bias, such as meta forecasts that pool multiple survey forecasts at time t to form a meta forecast, are ruled out because individual survey respondents do not have access to all the other analysts predictions in real time.

Third, since in principle all survey respondent-types could have accessed the same information given the machine, the time t machine forecast serves as a real-time check on whether the survey response may be making a demonstrable systematic expectational error. In practice, this artificial intelligence approach could be employed within institutions to check for, and possibly correct, biases in professional forecasts.

Below we compare the forecast accuracy of the machine benchmark with the survey responses. If the machine systematically improves forecasts on average over an extended evaluation sample, we take that as evidence of belief distortion, or "bias" for short. In this event, we compute a *dynamic* measure of a survey respondent-type's belief distortion by taking the difference between the survey forecast and the machine forecast, $\mathbb{E}_t^{(i)}[y_{j,t+h}]$, where we denote the bias of forecaster i at time t as

$$bias_{j,t}^{(i)} \equiv \mathbb{F}_t^{(i)}[y_{j,t+h}] - \mathbb{E}_t^{(i)}[y_{j,t+h}]. \tag{3}$$

Observe that $bias_{j,t}^{(i)}$ captures *ex ante* expectational errors, not *ex post* forecast errors, or "mistakes." In particular, bias in expectations is measured relative to the machine forecast, not relative to an *ex post* outcome. One implication of this is that it is possible that every respondent-type is biased vis-a-vis the machine *ex ante*, even though there will always be some respondent-type that is "right" *ex post*.

3.3 Private and Public Signals

To facilitate an interpretation of the machine estimates of $\alpha_j^{(i)}$, $\beta_{j\mathbb{F}}^{(i)}$, and $\mathbf{B}_{j\mathcal{Z}}^{(i)}$, we consider a framework in which forecasters form an overall prediction by combining a statistical forecast based on public information (i.e., a public signal) with a judgemental component based on

[5] See Table A.12 of the Online Appendix.

information intangible to the machine (i.e., a private signal). Let x be publicly available information and let z be a private signal about an unknown variable y. Suppose these variables are related to one another according to the system

$$\begin{aligned} x &\sim iid\left(0, \sigma_x^2\right) \\ y &= \alpha x + u_2,\ u_2 \sim N\left(0, \sigma_2^2\right) \\ z &= y + u_1,\ u_1 \sim N\left(0, \sigma_1^2\right), \end{aligned} \tag{4}$$

where α is a parameter describing the mapping from x to y, and where x, u_1, and u_2 are i.i.d. and mutually uncorrelated with one another. In what follows, we will interpret y as the future value of a variable being forecast, z as a private signal representing judgement or intangible information that a survey forecaster can observe but the machine cannot, and x as public information that serves both as a predictor and, via the mapping αx, as a public signal. Note that x could also be a vector, while αx is still a scalar. For example, if y is future inflation, x could be a measure of the output gap, the central bank inflation target, and/or macro and financial factors formed from large datasets. The random variable u_2 is the unforecastable component of y and has the interpretation of a structural shock.

Conditional on observing both the private and public signals, the optimal forecast of y is

$$\mathbb{E}_o\left[y|\alpha x, z\right] = \gamma z + (1-\gamma)\,\alpha x, \qquad \gamma \equiv \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} = \frac{\sigma_1^{-2}/\sigma_2^{-2}}{1 + \sigma_1^{-2}/\sigma_2^{-2}},$$

where $\sigma_1^{-2}/\sigma_2^{-2}$ is the precision of the private signal relative to that of the public signal. More weight is given to the private signal when it is relatively more precise.

Suppose that a forecaster assigns weights to her private and public signals as follows:

$$\mathbb{F} = \gamma^F z + \left(1 - \gamma^F\right)\alpha^F x.$$

The survey forecast $\mathbb{F}$ is here interpreted as a prediction based partly on the respondent's statistical model using public information ($\alpha^F x$) and partly on a private signal z. The case $\alpha^F \neq \alpha$ arises when the forecaster inefficiently combines the public information when forecasting y, which could occur if she is inattentive to some variables and/or overly attentive to others.

How might γ^F be determined? One possibility is that the forecaster chooses $\mathbb{F}$ based on what would be optimal given the system (4), but employs incorrect estimates $\sigma_{1F}^2 \neq \sigma_1^2$, $\alpha^F \neq \alpha$, in which case we have

$$\gamma^F = \frac{\sigma_{1F}^{-2}/\sigma_{2F}^{-2}}{1 + \sigma_{1F}^{-2}/\sigma_{2F}^{-2}},$$

where $\sigma_{1F}^{-2}/\sigma_{2F}^{-2}$ is the ratio of her *estimated* private signal precision to her *estimated* public signal precision. Note that if the forecaster inefficiently combines the public information, i.e., $\alpha^F \neq \alpha$, she will have effectively under-estimated the precision of the public signal, leading to $\sigma_{2F}^{-2} < \sigma_2^{-2}$. At the same time, it is possible that $\gamma^F = \gamma$ even if $\alpha^F \neq \alpha$ if the forecaster under-estimates the precision of her private signal by exactly the right amount.

Now consider the machine forecast $\mathbb{E}$ of y, which is based on both the public information x and the survey forecast $\mathbb{F}$:

$$\mathbb{E} = \widehat{\beta}\mathbb{F} + \widehat{B}x = \widehat{\beta}\left[\gamma^F z + \left(1 - \gamma^F\right)\alpha^F x\right] + \widehat{B}x,$$

where $\widehat{\beta}$ and $\widehat{B}$ are coefficient estimates. Although the machine cannot directly observe the private signal z, it can still learn about the weight assigned to it by the forecaster from observing $\mathbb{F}$. The machine estimates the coefficients $b \equiv (\beta, B)'$ from a regression of y on $(\mathbb{F}, x)'$. The Online Appendix proves that this estimator results in the values:

$$\widehat{\beta} = \frac{\gamma}{\gamma^F}, \quad \widehat{B} = (1-\gamma)\alpha - \left(\frac{\gamma}{\gamma^F} - \gamma\right)\alpha^F.$$

The machine will set $\widehat{\beta} < 1$ ($\widehat{\beta} > 1$) if and only if the forecaster gives more (less) weight γ^F to her private signal z than the correct weight based on its true relative precision $\sigma_1^{-2}/\sigma_2^{-2}$. Thus, the machine corrects for inefficiencies in the survey forecast by setting $\widehat{\beta} \neq 1$ and/or $\widehat{B} \neq 0$.

In the above framework, the case of $\gamma^F/\gamma > 1$ resulting in $\widehat{\beta} < 1$, could happen for two reasons. First, the forecaster might over-estimate the precision of her private signal, i.e., $\sigma_{1F}^{-2} > \sigma_1^{-2}$, a circumstance often referred to as "overconfidence" in the behavioral economics literature. Second, the forecaster might inefficiently combine the public information, i.e., $\alpha^F \neq \alpha$, so that she effectively under-estimates the precision of her public signal, i.e., $\sigma_{2F}^{-2} < \sigma_2^{-2}$. Either way, $\widehat{\beta} < 1$ indicates an overreliance, in relative terms, on the private or judgemental component of her forecast.[6] We return to a discussion of this case below when we present the machine parameter estimates.

3.4 Estimator

We now describe the machine estimation used to quantify any belief distortions. To do so, let us simplify notation by collecting all the independent variables and coefficients on the right-hand-side of (1) into a single matrix and vector and writing the machine model as:

$$y_{j,t+h} = \mathcal{X}_t' \boldsymbol{\beta}_j^{(i)} + \epsilon_{jt+h} \tag{5}$$

where $\mathcal{X}_t = \left(1, \mathbb{F}_t^{(i)}\left[y_{j,t+h}\right], \mathcal{Z}_{jt}\right)'$ and $\boldsymbol{\beta}_j^{(i)} \equiv \left(\alpha_j^{(i)}, \beta_{j\mathbb{F}}^{(i)}, \left(\mathbf{B}_{j\mathcal{Z}}^{(i)}\right)\right)'$.

[6]In principle $\hat{\beta} < 1$ could also occur if there is idiosyncratic measurement error in the survey responses. While possible, we argue that this is unlikely to be a plausible explanation for the findings reported below, for two reasons. First, such an interpretation is implausible for professional forecasters where $\hat{\beta}_{j\mathbb{F},t}^{(i)}$ is nonetheless quite substantially below unity on average. Second, measurement errors should wash out in the mean survey forecast, yet estimates of this parameter for the mean, and of the average bias, are similar to those for the median.

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

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

where $m\left(\mathbf{X}_T, \boldsymbol{\lambda}^{(i)}\right)$ defines an estimator as a function of the data $\mathbf{X}_T$ and a non-negative regularization or "tuning" parameter vector $\boldsymbol{\lambda}^{(i)}$ estimated using cross-validation. The values of $\boldsymbol{\lambda}^{(i)}$, which will be estimated dynamically over time, determine the optimal shrinkage and sparsity of the time t machine specification. Denote this latter estimator $\widehat{\boldsymbol{\lambda}}^{(i)}$ and denote the combined final estimator $\hat{\boldsymbol{\beta}}_j^{(i)}\left(\mathbf{X}_T, \widehat{\boldsymbol{\lambda}}^{(i)}\right)$. Our main approach uses the elastic net (EN) estimator, where $\boldsymbol{\lambda}^{(i)}$ is a bivariate vector that uses dual lasso and ridge penalties to achieve both shrinkage and sparsity.[7]

The estimation of (5) is repeated sequentially in rolling subsamples, with parameters estimated from information known at time t used to predict variables $y_{j,t+h}$ in *subsequent* periods. This leads to a sequence of machine efficient beliefs about $y_{j,t+h}$. Denote the coefficients and regularization parameters obtained from an estimation conducted with information through time t as $\hat{\boldsymbol{\beta}}_{j,t}^{(i)}$ and $\hat{\boldsymbol{\lambda}}_t^{(i)}$, respectively. Note that the time t subscripts on $\hat{\boldsymbol{\beta}}_{j,t}^{(i)}$ and $\widehat{\boldsymbol{\lambda}}_t^{(i)}$ are used to denote one in a sequence of time-invariant parameter estimates obtained from rolling subsamples, rather than estimates that vary over time within a sample. Likewise, we shall denote the time t machine belief about $y_{j,t+h}$ as $\mathbb{E}_t^{(i)}[y_{j,t+h}]$, defined by

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

Forecast errors are differentially denoted for the survey and machine

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

Survey and machine MSEs denoted with $\mathbb{F}$ and $\mathbb{E}$ subscripts, i.e.,

$$\text{survey MSE} \equiv MSE_{\mathbb{F}} = (1/P)\textstyle\sum_{i=1}^{P}(\text{survey error}_{t+h}^{(i)})^2 \qquad (6)$$

$$\text{machine MSE} \equiv MSE_{\mathbb{E}} = (1/P)\textstyle\sum_{i=1}^{P}(\text{machine error}_{t+h}^{(i)})^2 \qquad (7)$$

where P is the length of the forecast evaluation sample. To reduce notation clutter, we leave off superscripts "(i)" in the definitions above, but the reader is reminded that these statistics also depend on the respondent-type. Distortions in survey responses are quantified by the ratio $MSE_{\mathbb{E}}/MSE_{\mathbb{F}}$ over an extended forecast evaluation sample of size P. To uncover any belief distortions, the machine needs to solve a high-dimensional, dynamic, out-of-sample learning problem. We discuss our algorithm for doing so next.

[7] We have also implemented the approach in simulated data and hold-out samples for lasso and ridge separately, for random forest, and for empirical Bayes linear regression. The EN estimator was the best performing, followed by lasso, while random forest and Bayesian regression performed poorly.

3.5 Machine Learning Algorithm

This section discusses a novel machine learning algorithm developed to detect demonstrable, *ex ante* expectational errors in real time. This algorithm is explicitly designed to combat overfitting and cope with structural change in a dynamic setting. 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 with the EN estimator using observations $1, ..., T_{IS}$, given an arbitrary fixed (non-random) starting value for $\boldsymbol{\lambda}_t^{(i)}$. Denote this initial estimate $\boldsymbol{\beta}_{T_{IS}}^{*(i)}\left(\mathbf{X}_{T_{IS}}, \boldsymbol{\lambda}_t^{(i)}\right)$, where "*" denotes the value of the estimator given an arbitrary $\boldsymbol{\lambda}_t^{(i)}$.
3. **Training and cross-validation**: The regularization parameter $\boldsymbol{\lambda}_t^{(i)}$ is estimated by minimizing mean-square loss $\mathcal{L}\left(\boldsymbol{\lambda}_t^{(i)}, T_{IS}, T_{TS}\right)$ over *pseudo* out-of-sample forecast errors generated from rolling regressions through the training sample, where

$$\mathcal{L}\left(\boldsymbol{\lambda}_t^{(i)}, T_{IS}, T_{TS}\right) \equiv \frac{1}{T_{TS}-h} \sum_{\tau=T_{IS}}^{T_{IS}+T_{TS}-h} \left(\mathcal{X}_\tau' \boldsymbol{\beta}_{j,\tau}^{*(i)}\left(\mathbf{X}_{T_{IS}}, \boldsymbol{\lambda}_t^{(i)}\right) - y_{j,\tau+h}\right)^2, \qquad (8)$$

   and where $\boldsymbol{\beta}_{j,\tau}^{*(i)}\left(\mathbf{X}_{T_{IS}}, \boldsymbol{\lambda}_t^{(i)}\right)$ is the time τ EN estimate of $\boldsymbol{\beta}_j^{(i)}$ given $\boldsymbol{\lambda}_t^{(i)}$ 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 estimator of $\boldsymbol{\beta}_{j,t}^{(i)}\left(\mathbf{X}_{T_{IS}}, \boldsymbol{\lambda}_t^{(i)}\right)$ is based on the most recent $\widehat{T}_{IS}$ observations where $\left\{\widehat{\boldsymbol{\lambda}}_t^{(i)}, \widehat{T}_{IS}, \widehat{T}_{TS}\right\} = \underset{\boldsymbol{\lambda}, T_{IS}^*, T_{TS}^*}{\operatorname{argmin}} \mathcal{L}\left(\boldsymbol{\lambda}_t^{(i)}, T_{IS}^*, T_{TS}^*\right)$ and is denoted $\widehat{\boldsymbol{\beta}}_{j,t}^{(i)}\left(\mathbf{X}_{\widehat{T}_{IS}}, \widehat{\boldsymbol{\lambda}}_t^{(i)}\right)$.
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 $\widehat{\boldsymbol{\beta}}_{j,t}^{(i)}\left(\mathbf{X}_{\widehat{T}_{IS}}, \widehat{\boldsymbol{\lambda}}_t^{(i)}\right)$ and the machine forecast error $y_{t+h} - \mathcal{X}_t' \widehat{\boldsymbol{\beta}}_{j,t}^{(i)}\left(\mathbf{X}_{\widehat{T}_{IS}}, \widehat{\boldsymbol{\lambda}}_t^{(i)}\right)$ stored.
6. **Roll forward and repeat**: The prior sample of data is rolled forward one period, and steps 1-5 are repeated. This continues until the last out-of-sample forecast is made for $y_{j,T}$, where T is the last period of our sample.

Referring back to the notation in (6) and (7), $MSE_{\mathbb{E}}$ is computed by averaging across the sequence of squared forecast errors in the true out-of-sample forecasting step 5 for periods $t = (\widetilde{T} + h)$,...,T. We refer to this subperiod as the external forecast *evaluation sample*.

Several points about the above procedure bear emphasizing. First, the algorithm ensures that the machine forecast selected from step 4 can only differ from the survey forecast if it demonstrably improves pseudo out-of-sample prediction in the rolling training samples *prior* to making a true out-of-sample forecast in step 5. Otherwise, the machine adopts the survey forecast. It follows that the true out-of-sample forecasts of the machine recorded in step 5 can differ from those of the survey only if demonstrable, *ex ante* biases are detected. The resulting measure of belief distortion therefore explicitly excludes *ex post mistakes* that the machine algorithm could only have understood with hindsight. An implication of this *ex ante* approach is that more than one type can show no bias if the machine is unable to detect patterns in extraneous economic data that can be exploited in real time to improve forecasts. We quantify the overall magnitude of forecaster bias with the ratio $MSE_{\mathbb{E}}/MSE_{\mathbb{F}}$ taken over the evaluation sample.

Second, the machine algorithm is repeated for each i and for each t in the evaluation sample. This can be important for all the parameter estimates but especially so for the estimate of the intercept, which functions as a latent time-varying mean.

Third, each new training renews the optimized selection of in-sample estimation and training sample windows lengths, innovating on traditional machine learning approaches by paying close attention to the time structure of data. This is important because, in a dynamic setting subject to possible structural change, no single set of window lengths is likely to work best in all time periods.[8] The algorithm developed here instead asks the machine to choose both the optimal estimation window and the optimal window for determining shrinkage and sparsity, dynamically. This may be contrasted with traditional "K-fold" cross-validation techniques, which partition a sample in arbitrary ways, leading to tuning parameters that are chosen partly on the basis of how well the future predicts the past. Table (A.2) of the Online Appendix shows that the machine performs poorly when it is trained using traditional K-fold cross-validation techniques.

Fourth, the cumulation of true out-of-sample forecast errors from step 5 serves as an external validation step that exists outside of the optimization loop. It is crucial that the number of forecast error observations aggregated from this step be large enough so that the evidence on relative forecast accuracy is not the result of a few random outliers. For this reason, we require that our external evaluation samples from this step be at least 84 quarters long for all machine specifications. At the same time, we must balance this imperative against two others: the need to reserve a minimum number of observations to do estimation and training, and, because there are differences in data availability across the surveys, the need to compare relative forecast using roughly comparable time frames. For the SPF, data on inflation and GDP growth forecasts are available from 1969:Q3 to 2018:Q3. Thus the machine can partition a prior sample up to size $\widetilde{T} = 98$ quarters in every step of the recursion and still produce an external evaluation

[8] See Giacomini and White (2006) and Pesaran and Timmermann (2007) for extensive evidence related to these themes.

sample with 97 quarterly observations that spans 1995:Q1 to 2018:Q2. In this case, the first true out-of-sample forecast of four-quarter-ahead outcomes is recorded for the period 1994:Q1 to 1995:Q1. For SOC, both inflation and GDP growth forecast data are available from 1978:Q1 to 2018:Q2, while for BC, inflation forecasts are available from 1986:Q1 to 2018:Q2, and GDP forecasts are available from 1984:Q3 to 2018:Q2. Since these surveys extend less far backward in time, estimation and training must be accomplished on smaller prior sample sizes $\widetilde{T}$ in the early recursions in order to ensure an external evaluation sample of at least 84 observations. But it is still possible to allow for reasonable minimal prior sample sizes at each step in these recursions, while nonetheless ending up with external evaluation samples of at least 84 quarters that cover roughly comparable time periods. Our external evaluation sample for the SOC surveys consists of 97 quarterly observations and span 1995:Q1 to 2018:Q2. For the BC surveys the external evaluation sample for inflation forecasts consists of 84 quarterly observations and spans 1997:Q3 to 2018:Q2, while that for the GDP growth forecasts consists of 89 quarterly observations and spans 1996:Q1-2018:Q2.

3.6 Switching Model

When forecasting GDP growth, it is important to account for possible turning points in the economy. Since at least the ground breaking contribution of Hamilton (1989), it is well known that aggregate output growth is well described by a process that evolves differently across distinct economic states associated with recessions and expansions. A large subsequent literature treats models with regime changes as part of the standard forecasting toolbox for output growth (e.g., Chauvet and Potter (2013)). Furthermore, by the mid-1990s, a large body of evidence had accumulated that the slope of the term structure of interest rates had strong predictive power for the US business cycle. Specifically, inversions or a flattening of the yield curve typically anticipate a recession.[9] (By contrast, there was much less evidence that term spreads were especially useful in forecasting inflation at any time in the business cycle.)

Putting this all together, a forecaster operating in the mid-1990s would have had access to a large body of evidence indicating that (i) output growth behaves differently in recessions than in expansions, and (ii) turning points are often anticipated by a flat or inverted yield curve. Based on this prior knowledge, the machine efficient benchmark is specified to follow a simple switching model for output growth (but not inflation), for all forecasts starting in 1995:Q1.

To implement this idea, we combine the notion of distinct regimes with the predictive power of the term structure slope using a threshold model. This feature allows the machine to choose in real time whether to switch to a simpler, recession specification. The threshold aspect works

[9] See Harvey (1989), Estrella and Hardouvelis (1990, 1991), Plosser and Rouwenhorst (1994), Haubrich and Dombrosky (1996), Kozicki et al. (1997), Dotsey (1998) and Estrella and Mishkin (1998).

as follows:

$$
\begin{aligned}
y_{j,t+h} &= \alpha_j^{(i)} + \beta_{j\mathbb{F}}^{(i)} \mathbb{F}_t^{(i)} \left[y_{j,t+h} \right] + \mathbf{B}_{j\mathcal{Z}}^{(i)\prime} \mathcal{Z}_{jt} + \epsilon_{jt+h} \quad && \text{if } slope_{kt} > \widehat{tr}_{kt} \\
y_{j,t+h} &= B_{kt} I_{kt} && \text{if } slope_{kt} \leq \widehat{tr}_{kt},
\end{aligned}
$$

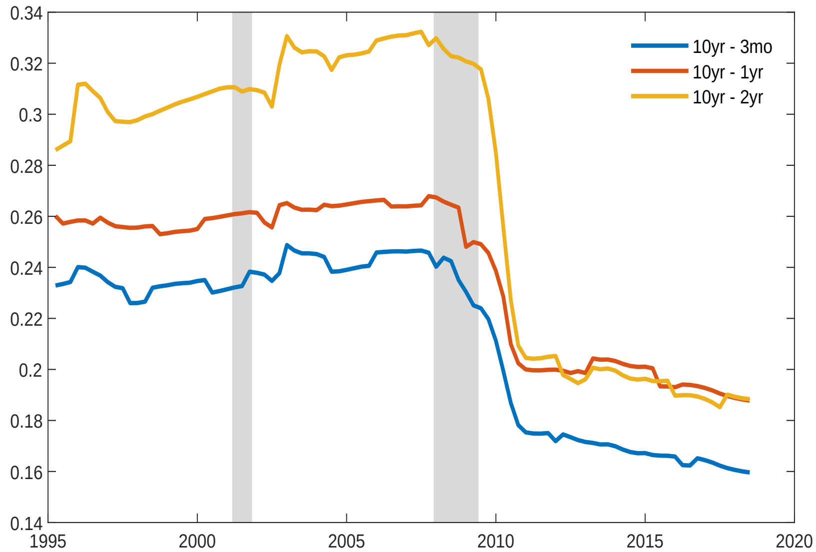
where B_{kt} is a parameter and I_{kt} is a dummy variable that depends on a yield curve measure at time t.

At each point in time, the machine chooses whether to use the "normal-times" specification (first row) or the "recession" specification (second row). The normal-times specification is based on the machine algorithm discussed in the previous section. The recession specification is chosen whenever $slope_{kt} < \widehat{tr}_{kt}$, where $slope_{kt}$ is a yield spread measure at time t and $\widehat{tr}_{kt}$ is a threshold. We consider three different yield curve slope indicators, indexed by k: the 10yr-3mo Treasury spread ($10y3m$), the 10yr-1yr Treasury spread ($10y1y$), and the 10yr-2yr Treasury spread ($10y2y$). The machine uses past data to run a regression of GDP growth h quarters ahead (Δgdp_{t+h}) on a dummy variable I_{kt} that equals 1 when $slope_{kt} \leq tr_{kt}$ and 0 otherwise. The machine searches in real time for the specific threshold $\widehat{tr}_{kt}$ and the yield spread indicator $slope_{kt}$ that maximizes the R^2 of this regression. The machine repeatedly re-optimizes the choice of both $\widehat{tr}_{kt}$, and the specific measure of the term spread ($10y3m$, $10y1y$, or $10y2y$) based on real-time forecasting regressions of GDP growth on I_{kt} using expanding windows of data up to time t and beginning in 1976:Q3, when the data on the two-year Treasury bill rate is first available. The time t recession specification forecast of GDP growth in $t+h$ is then simply the average GDP growth over all periods where $I_t = 1$ in a sample spanning 1976:Q3 to t.

Figure 2 reports the real-time R^2s from the expanding-window regressions of GDP growth at $t+h$ on I_{kt} using the different measures of the term spread. The figure shows that regressions using the $10y2y$ dummies are almost always chosen by the machine because that specification delivers the highest real-time predictive power for GDP growth in almost all periods of our evaluation sample, including those just prior to the two recessions in the sample that occurred in 2001 and 2007-2009.

It is worth noting, however, that the only time over the evaluation samples that the recession specification is triggered is just prior to the 2001 recession. In particular, there is no switch triggered prior to the Great Recession. This happens because all term spreads exhibited a secular decline between the two recessions, so that by 2007 the threshold values commensurate with a forecast of negative economic growth had also declined. At the same time, the declines in yield spreads prior to the 2007-2009 recession were relatively modest by historical standards and thus never fell below the (lower) real-time $\widehat{tr}_{kt}$ thresholds. In contrast to the 2001 recession, yield spreads generally failed to signal the Great Recession, a structural shift that shows up in Figure 2 as a sharp drop in real-time R^2 statistics right after the recession.

Figure 2: Real-time R^2 with Optimal Thresholds



This figure reports the R^2s from expanding-window regressions of GDP growth at t on a dummy variable I_{t-h}. The dummy variable I_t equals one if the term spread is below its real-time N_tth percentile at time t. The threshold N_t maximize the R^2 from the regression estimated at time t. The real time percentile of the term spread at time t is computed using the data from 1976:Q3 to time t. The yellow line reports the results using term spread defined as 10-year Treasury bond rate minus 2-year rate. The red line uses 10-year rate minus 1-year rate. The blue line uses 10-year rate minus 3-month rate. NBER recessions are shown with grey shaded bars. The sample is 1995:Q1-2018:Q2.

4 Data

The data used for this study fall into several categories. For each category the sources and details are left to the Appendix.

Survey Data The first data category is the survey data.

The SPF is a quarterly survey. Respondents provide both nowcasts and quarterly forecasts from one to four quarters ahead. We focus on the survey questions about the level of the GDP deflator (PGDP) and the level of real GDP. We use these data to construct forecasts of GDP growth, as explained in the Appendix. We also use SPF forecasts of 10-year-ahead CPI inflation as information variables.

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, and by how much, 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, we use questions in which forecasters are asked to predict the average quarter over quarter percentage change in Real GDP and the GDP deflator, beginning with the current quarter and extending four to five quarters into the future.

For all surveys, we align the timing of survey response deadlines with real-time data, so that the machine can only use data available in real time before the survey deadline.

Real-time Macro Data A real-time macro dataset provides observations on the left-hand-side variables on which forecasts are formed obtained from the Philadelphia Fed's Real-Time Dataset. Following Coibion and Gorodnichenko (2015), to construct forecasts and forecast errors, we use the vintage of inflation and GDP growth data that is available four quarters after the period being forecast. We also use the real-time macro data to form real-time quarterly macro factors from a constructed dataset of real-time quarterly macro variables observed on or before the day of the survey deadline at each date t. The resulting real-time macro dataset, contains observations on 92 real-time macro variables. Our real time macro variable dataset also include data on home and energy prices, which are not revised and so do not have multiple vintages. The complete list of macro variables is given in the Online Appendix.

Monthly Financial Data To take into account financial market data, we form factors from a panel dataset of 147 monthly financial indicators 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. We convert the monthly factors formed from the dataset into quarterly factors by using the first month's observation for each quarter.

Daily Financial Data "Up-to-the-forecast" financial market information is accounted for by using daily data on financial indicators up to one day before the survey respondents forecasts are due. The daily financial dataset includes series from five broad classes of financial assets: (i) commodities prices (ii) corporate risk variables including a number of different credit spreads measuring default risk (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), with the complete set of variables reported in the Online Appendix. In order to use both daily and quarterly data in our estimation, we combine diffusion index estimation of daily financial factors with mixed data sampling frequency techniques, described in detail in the Appendix.

Additional Non-Factor Data A number of other non-factor variables are also included 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 forecasts and those of other percentiles, higher-order cross-sectional moments of the lagged forecast distributions, several autoregressive lags of the left-hand-side variables, long-term trend inflation measures, and measures of detrended employment and GDP (Hamilton, 2018).

In all, once factors are formed the machine model entertains a total of 68 predictor variables for inflation and 72 predictor variables for GDP growth, before the machine chooses sparsity. Below we refer to estimated factors with an economic name. The economic name makes use of group classifications for individual series and output from time series regressions of individual series onto estimated factors, for each time period in our evaluation sample. For example, if regressions of non-farm payrolls onto the first common macro factor from the real-time macro panel dataset exhibits the highest average (across all time periods of our evaluation sample) marginal R^2, then that factor is labeled an "Employment" factor and normalized so that it increases when non-farm payrolls increase.

5 Results

This section reports results using our estimates of belief distortions across different respondent-types, surveys, and variables. In all cases, we focus on $h = 4$ quarter-ahead forecasts.

5.1 Forecast Comparison

We present a comparison of the accuracy of forecasts made by the machine benchmark and the survey respondents over the external evaluation samples. Table 1 reports the ratio of the machine $MSE_{\mathbb{E}}$ to the survey $MSE_{\mathbb{F}}$ for inflation and GDP growth for all three surveys over their respective external evaluation samples, along with several other results. We discuss these in turn.

First consider the average predictive accuracy of the machine versus the forecaster-type over the external evaluation sample. The top panel of Table 1 shows that the machine model performs better than the survey forecasts of inflation for all surveys and all respondent-types as measured by the ratio $MSE_{\mathbb{E}}/MSE_{\mathbb{F}}$, which is less than one in all cases, sometimes by large amounts. To put this ratio in the same units as an in-sample R^2, the table also reports an out-of-sample R^2 for the machine vis-a-vis the survey as $R^2_{OOS} \equiv 1 - MSE_{\mathbb{E}}/MSE_{\mathbb{F}}$. The overall magnitude by which the machine model improves on the survey forecasts is in most cases sizable, which is notable since survey forecasts of inflation are known to be difficult to beat or even match by statistical models out-of-sample, as discussed above. For example, the ratio $MSE_{\mathbb{E}}/MSE_{\mathbb{F}}$ for the median SPF forecast is 0.85. These ratios are similar for the median BC survey, as shown in the last panel, where in this case $MSE_{\mathbb{E}}/MSE_{\mathbb{F}}$ is 0.84. 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

Table 1: **Machine Learning versus Survey Forecasts**

ML: $y_{j,t+h} = \alpha_j^{(i)} + \beta_{j\mathbb{F}}^{(i)} \mathbb{F}_t^{(i)} [y_{j,t+h}] + \mathbf{B}_{j\mathcal{Z}}^{(i)} \mathcal{Z}_{jt} + \epsilon_{jt+h}$													
Inflation Forecasts													
Survey of Professional Forecasters (SPF)													
Percentile	Median	5th	10th	20th	25th	30th	40th	60th	70th	75th	80th	90th	95th
$MSE_{\mathbb{E}}/MSE_{\mathbb{F}}$	0.85	0.56	0.74	0.83	0.90	0.88	0.89	0.74	0.70	0.67	0.59	0.55	0.47
OOS R^2	0.15	0.44	0.26	0.17	0.10	0.12	0.11	0.26	0.30	0.33	0.41	0.45	0.53
w^*	0.68	0.82	0.78	0.74	0.63	0.66	0.66	0.78	0.76	0.74	0.80	0.79	0.83
$MSE_{\mathbb{R}}/MSE_{\mathbb{F}}$	0.82	0.41	0.64	0.77	0.81	0.82	0.84	0.76	0.66	0.61	0.53	0.39	0.27
Michigan Survey of Consumers (SOC)													
Percentile	Median	5th	10th	20th	25th	30th	40th	60th	70th	75th	80th	90th	95th
$MSE_{\mathbb{E}}/MSE_{\mathbb{F}}$	0.62	0.22	0.27	0.45	0.58	0.69	0.70	0.40	0.22	0.16	0.13	0.05	0.03
OOS R^2	0.38	0.78	0.73	0.55	0.42	0.31	0.30	0.60	0.78	0.84	0.87	0.95	0.97
w^*	1.00	0.96	0.93	0.92	0.92	0.95	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$MSE_{\mathbb{R}}/MSE_{\mathbb{F}}$	0.62	0.11	0.20	0.42	0.52	0.64	0.76	0.37	0.21	0.16	0.11	0.04	0.02
Blue Chip Financial Forecasts (BC)													
Percentile	Median	5th	10th	20th	25th	30th	40th	60th	70th	75th	80th	90th	95th
$MSE_{\mathbb{E}}/MSE_{\mathbb{F}}$	0.84	0.58	0.60	0.85	0.85	0.86	0.91	0.78	0.69	0.65	0.59	0.48	0.38
OOS R^2	0.16	0.42	0.40	0.15	0.15	0.14	0.09	0.22	0.31	0.35	0.41	0.52	0.62
w^*	0.65	0.73	0.76	0.62	0.63	0.62	0.58	0.70	0.79	0.82	0.86	0.94	0.92
$MSE_{\mathbb{R}}/MSE_{\mathbb{F}}$	0.76	0.45	0.55	0.72	0.76	0.78	0.78	0.73	0.66	0.62	0.57	0.43	0.33
GDP Forecasts													
Survey of Professional Forecasters (SPF)													
Percentile	Median	5th	10th	20th	25th	30th	40th	60th	70th	75th	80th	90th	95th
$MSE_{\mathbb{E}}/MSE_{\mathbb{F}}$	0.88	0.70	0.81	0.80	0.84	0.87	0.88	0.85	0.81	0.79	0.80	0.69	0.64
OOS R^2	0.12	0.30	0.19	0.20	0.16	0.13	0.12	0.15	0.19	0.21	0.20	0.31	0.36
w^*	0.84	0.96	1.00	1.00	1.00	0.94	0.87	0.85	0.88	0.87	0.83	0.85	0.84
$MSE_{\mathbb{R}}/MSE_{\mathbb{F}}$	0.87	0.74	0.83	0.88	0.89	0.89	0.88	0.84	0.81	0.79	0.77	0.67	0.58
Michigan Survey of Consumers (SOC)													
Percentile	Median												
$MSE_{\mathbb{E}}/MSE_{\mathbb{F}}$	0.78												
OOS R	0.22												
w^*	0.81												
Blue Chip Financial Forecasts (BC)													
Percentile	Median	5th	10th	20th	25th	30th	40th	60th	70th	75th	80th	90th	95th
$MSE_{\mathbb{E}}/MSE_{\mathbb{F}}$	0.78	0.77	0.74	0.89	0.83	0.82	0.78	0.78	0.79	0.75	0.71	0.67	0.67
OOS R^2	0.22	0.23	0.26	0.11	0.17	0.18	0.22	0.22	0.21	0.25	0.29	0.33	0.33
w^*	0.75	0.68	0.75	0.60	0.73	0.71	0.76	0.72	0.71	0.72	0.77	0.77	0.74
$MSE_{\mathbb{R}}/MSE_{\mathbb{F}}$	0.75	0.70	0.76	0.79	0.79	0.78	0.76	0.73	0.70	0.69	0.66	0.60	0.55

Machine v.s. survey mean-square-forecast errors. $\text{MSE}_{\mathbb{E}}$ and $\text{MSE}_{\mathbb{F}}$ denote the machine and survey mean-squared-forecast-errors, respectively, for 4-quarter-ahead forecasts, averaged over the evaluation sample. The out-of-sample Rsquared, OOS R^2, is defined as 1-$\text{MSE}_{\mathbb{E}}/\text{MSE}_{\mathbb{F}}$. w^* denotes the weight on the machine forecast for the hybrid forecast described in the text. $\text{MSE}_{\mathbb{R}}$ denotes the MSE from FIRE benchmark described in the text. The vintage of observations on the variable being forecast is the one available four quarters after the period being forecast. The evaluation period for the Survey of Professional Forecasters (SPF) and the Michigan Survey of Consumers (SOC) is 1995:Q1 to 2018:Q2; and for the Bluechip (BC) survey is 1997:Q3 to 2018:Q2 (inflation) and 1996:Q1 to 2018:Q2 (GDP).

median $MSE_{\mathbb{E}}/MSE_{\mathbb{F}}$ ratio is 0.42, respectively, implying large out-of-sample R^2 statistics.[10]

For GDP growth, the lower panel of Table 1 shows that machine model is again always more accurate than the survey respondent-type no matter which respondent-type or survey is studied. The $MSE_{\mathbb{E}}/MSE_{\mathbb{F}}$ ratios for the median SPF and BC forecasts of GDP growth are 0.89 and 0.76, respectively. For the SOC, there is only a single forecast, denoted as if it corresponds to the "median" household, since the SOC forecast is constructed from the balance score for business conditions expectations, eliminating the heterogeneity (see above). The $MSE_{\mathbb{E}}/MSE_{\mathbb{F}}$ for this single SOC forecast of GDP growth is 0.74. In the next section we show that the gains in forecast accuracy afforded by the machine are even greater in the last five years of our sample, from 2013:Q2 to 2018:Q2.

While a comparison of mean squared forecast errors is one sensible way to compare predictive accuracy, there are obviously others. For example, frequentist econometric tests are sometimes employed to compare the predictive accuracy of two models in statistical terms. While this is certainly reasonable for some questions, we argue that such tests are unlikely to be especially useful or relevant when the objective is to measure belief distortions in survey point forecasts, as here. Consider that a professional forecaster might have two models that produce economically large differences in mean squared forecast error, while statistical tests can often indicate that they are the "same." The question then becomes, what should a forecaster do with that information? After all, survey respondents must still choose their best point prediction. If mean square loss is the objective, there is only one optimal choice, and this is unaffected by the amount of sampling noise around her two model forecasts. A further limitation of such tests is that they return merely a binary answer on whether a null hypothesis is rejected or not, while being silent on the practical quantitative question of by how much one model is more accurate than another.

For these reasons we take an alternative approach to model comparison, one that allows us to quantify any gains in forecast accuracy on a continuum from low to high, without being affected by the overall amount of statistical noise in the environment that both the machine and the survey forecast are subject to. The approach is motivated by the work of Amisano and Geweke (2017), who consider the properties of weighted linear combinations of prediction models. The key idea is that even if one model has superior predictive power over others, an optimal linear combination typically includes several models with positive weights, since being better on average is not synonymous with being always better. Amisano and Geweke (2017) focus on density forecasts, but given that survey responses are point forecasts rather than density forecasts, we adapt their idea by solving for the optimal linear combination of the

[10]Table A.1 in the Online Appendix shows that the machine also improves over the mean of the survey forecasts. We do not report those results here because the mean is always an amalgam that does not correspond to the belief of any single respondent-type in the survey and would not be known to any individual. It is arguably less relevant to the study of what, if any, systematic errors individuals may make when forming macroeconomic expectations.

machine and survey forecasts that minimizes the mean square forecast error over our evaluation sample. We refer to this linear combination as the optimal "hybrid" forecast.

Specifically, consider a hybrid forecast of $y_{j,t+h}$, denoted $\mathbb{EF}_t^{(i)}[y_{j,t+h}]$, obtained as a weighted average of the machine and the survey forecasts:

$$\mathbb{EF}_t^{(i)}[y_{j,t+h}] \equiv w\mathbb{E}_t^{(i)}[y_{j,t+h}] + (1-w)\,\mathbb{F}_t^{(i)}[y_{j,t+h}]$$

where $w \in [0,1]$. Conceptually we can ask, given the average performance of these two forecasts over our sample, how much weight w would one want to place on one versus the other in a hybrid forecast if we faced an identical sample in the future? To answer this question, note that the hybrid forecast errors are a linear combination of the machine and survey forecast errors:

$$\begin{aligned}
\text{hybrid error}_{t+h}^{(i)} &= \mathbb{EF}_t^{(i)}[y_{j,t+h}] - y_{j,t+h} \\
&= w\mathbb{E}_t^{(i)}[y_{j,t+h}] + (1-w)\,\mathbb{F}_t^{(i)}[y_{j,t+h}] - y_{j,t+h} \\
&= w(\text{machine error}_{t+h}^{(i)}) + (1-w)\,(\text{survey error}_{t+h}^{(i)})
\end{aligned}$$

with the hybrid mean square forecast error given by

$$\text{hybrid MSE} \equiv MSE_{\mathbb{EF}} = (1/P)\textstyle\sum_{i=1}^{P}(\text{hybrid error}_{t+h}^{(i)})^2.$$

The optimal weight w^* placed on the machine forecast is defined as the one that minimizes the hybrid MSE over our evaluation samples, i.e.,

$$w^* = \arg\min MSE_{\mathbb{EF}} = \arg\min\,(1/P)\textstyle\sum_{i=1}^{P}(\text{hybrid error}_{t+h}^{(i)})^2,$$

where P is the length of the evaluation sample.

The weights w^* are reported in the third row of each subpanel in Table 1. To interpret these numbers, note that if the machine were always better than the survey, w^* would be 1. This happens with many percentiles of the SOC inflation forecasts, and in several percentiles of the SPF GDP growth forecasts. If instead the machine is only marginally better than the survey, w^* would be close to 0.5. For the median GDP growth forecasts, the weights are well above 0.5, equal to 0.84, 0.81 and 0.81 for the SPF, SOC and BC median forecasts, respectively. The analogous weights for the median inflation forecasts are 0.68, 1, and 0.65, and typically higher for the other percentile types. These relatively large numbers close to unity imply that the machine produced economically meaningful gains in forecast accuracy over the survey responses during the historical sample over which the two forecasts were separately evaluated.

Finally, in the last rows of Table 1 we report the results of a different type of model comparison. Specifically, we compare the accuracy of survey forecasts with that from an alternative machine specification that differs from our baseline machine specification along only one dimension: it uses every percentile-type's time t forecast rather than just the ith percentile's. This alternative benchmark is motivated by certain full information, rational expectations (FIRE) models in which every agent in the economy receives a private signal, but other agents' private signals are not publicly known. In such models, the FIRE benchmark against which any distortion is measured is based on the union of everyone's information at time t.[11] The last rows of each panel in Table 1 report the ratio of the mean squared forecast errors under this FIRE benchmark, denoted $MSE_{\mathbb{R}}$, to the survey $MSE_{\mathbb{F}}$. Comparing the ratio $MSE_{\mathbb{R}}/MSE_{\mathbb{F}}$ with those in the first row showing the baseline $MSE_{\mathbb{E}}/MSE_{\mathbb{F}}$, we see that the numbers are quite similar, and in some cases the latter ratio is a bit smaller than the former. This shows that the improvement in forecast accuracy afforded by allowing the machine benchmark to observe everyone's time t prediction–where it exists at all–is minimal. This finding is relevant because it suggests that information frictions based on noisy "dispersed information" are unlikely to be the most relevant source of belief distortion in our data.

Returning to the baseline specifications, it is of interest to consider the nature of the empirical specifications chosen by the machine that produce gains in forecast accuracy. Figure 3 reports a scatter plot that quantifies the strength of the estimated ridge and lasso 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. We see that the machine algorithm often results in a sparse specification. In many time periods the fraction of non-zero coefficients hovers around 10% or less, though in some periods the machine chooses little if any sparsity, but much greater L^2 shrinkage. Occasionally, the machine chooses minimal sparsity and minimal L^2 shrinkage. This implies that achieving the efficiency gains of the machine over the extended evaluation sample requires entertaining large datasets in every period, even though much of that information is associated with a coefficient that is shrunk all the way to zero most of the time.

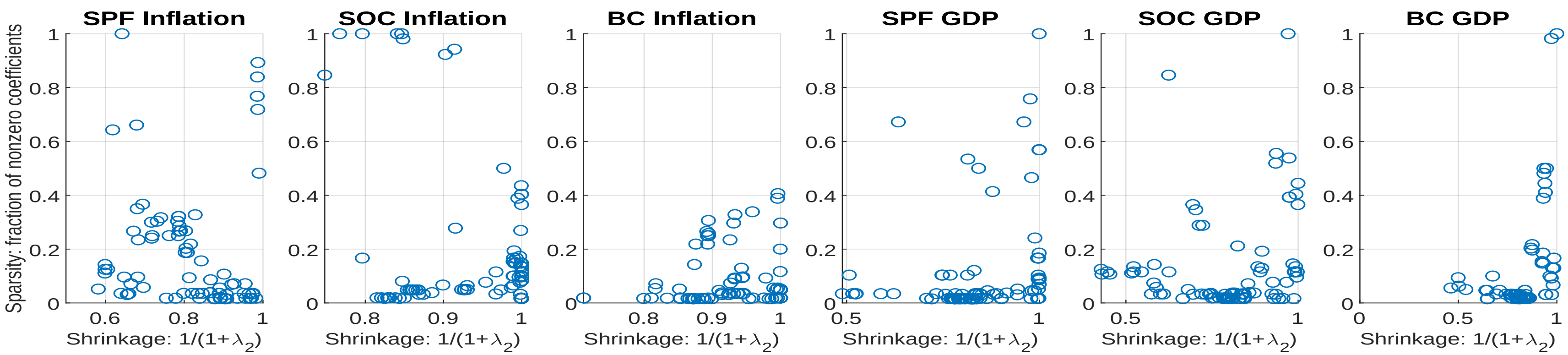
5.2 Dynamics of Belief Distortions

We now turn to investigate the dynamics of systematic expectational errors, by reporting the median bias over time, i.e., $bias_{j,t}^{(50)} \equiv \mathbb{F}_t^{(50)}\left[y_{j,t+h}\right] - \mathbb{E}_t^{(50)}\left[y_{j,t+h}\right]$ over our evaluation sample. Note that the units of $bias_{j,t}^{(50)}$ are the same as the forecasts themselves and are in annual percentage points. Figure 4 shows biases associated with the mean and median respondents for all three surveys.

Figure 4 shows that systematic errors in the median forecasts vary substantially over time and can range between 50% and 400% of the average annual inflation or GDP growth, depending on the survey. Survey forecasts for GDP growth oscillate between "optimism" and

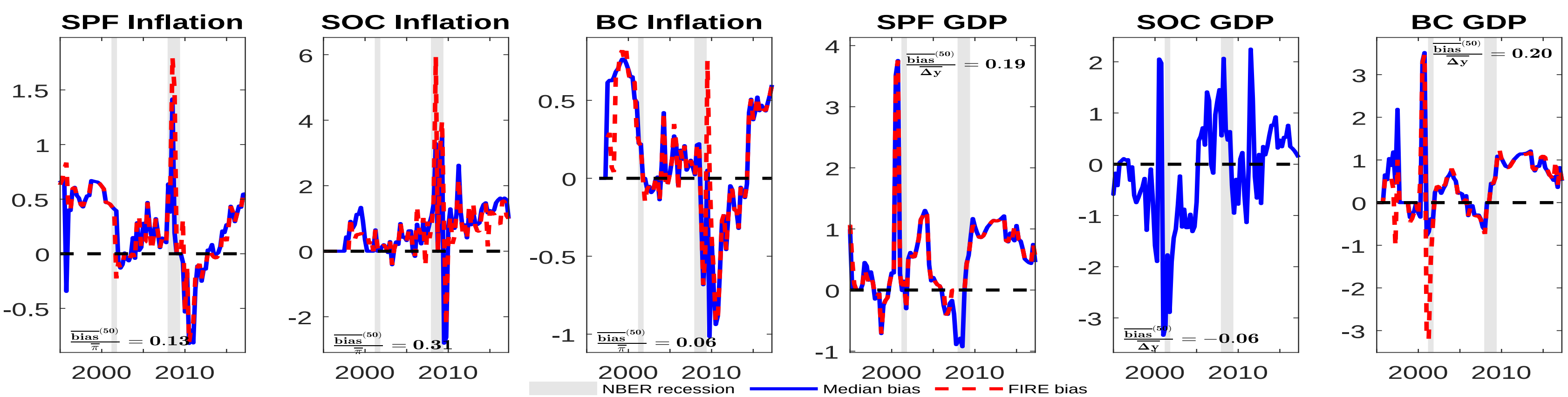
[11] We thank an anonymous referee for suggesting this alternative comparison.

Figure 3: 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 4: Biases in the Median Survey Forecasts



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

"pessimism." For GDP growth the figure shows extended periods of over-optimism that are especially prevalent for professional forecasters in the post-Great Recession part of our subsample. From 2010:Q1 to 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. This large upward bias since 2010 contributes heftily to the upward bias over the full evaluation sample (1995:Q1-2018:Q2), which is also sizable and amounts to 20% of observed GDP growth. These distortions are quite similar for the median BC expectation of GDP growth. For the SOC, the *average* bias is close to zero even though the SOC forecast is less accurate than the SPF or BC forecasts. This happens because the SOC forecast makes systematic errors of greater magnitude that fluctuate more wildly between optimism and pessimism. For all surveys, there are large spikes in the biases at the cusp of the 2000-2001 recession, a finding we discuss further below. Figure 4 also shows the biases implied by comparing to the median forecasts to the machine FIRE benchmark discussed above, which uses the union of everyone's information at time t, in addition to the extensive public information (red dashed line). With the exception of a few outlier observations, deviations of the median forecast from this FIRE benchmark track closely $bias_{j,t}^{(50)}$, again suggesting that noisy, dispersed information is not the most important driver of our measured belief distortions.

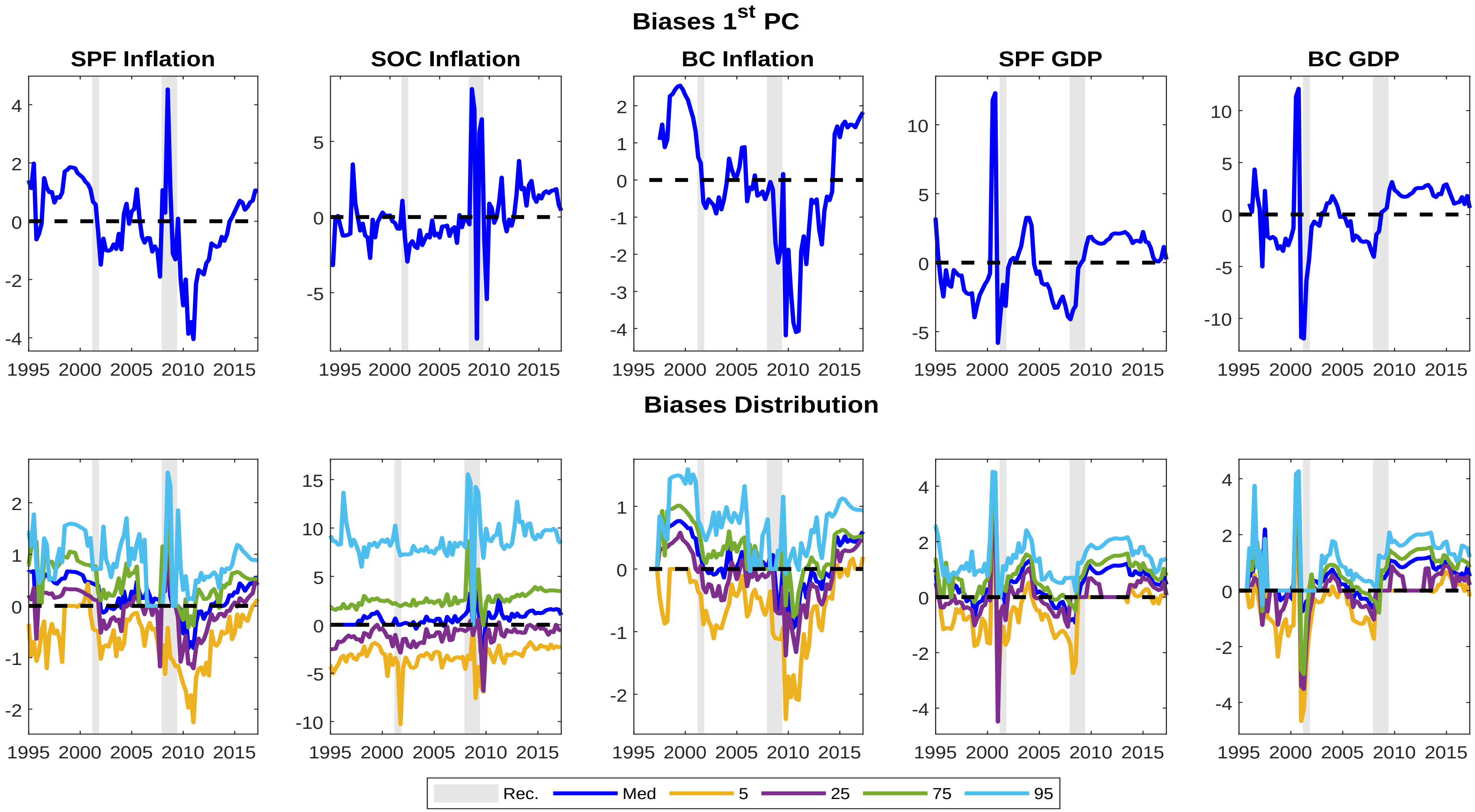
For inflation, Figure 4 shows that the median expectations are biased upward (a direction we defined above as "pessimistic") over most of the sample for the SPF and the SOC, while the BC survey exhibits an average bias that is close to zero.[12] Despite being upwardly biased on average over the full sample, median inflation forecasts exhibit a downward bias from 2011 to 2014 that ranges across surveys 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.

Figure 5 contrasts 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 (PC) of $bias_{j,t}^{(i)}$ across all percentiles i, with heterogeneity exhibited by the distribution of $bias_{j,t}^{(i)}$ across i.[13] For all surveys, we observe substantial variation in belief distortions over time that is common across SPF respondents. For SPF and BC, the optimism about economic growth in the immediate aftermath of the Great Recession is present in the

[12]Whether an upward bias in inflation expectations should be viewed as pessimism or optimism may depend on the time period. Bhandari, Borovicka, and Ho (2019) argue that a general interpretation of higher expected inflation as optimism is at odds with surveys of inflation attitudes, but others have argued that a downward bias in inflation expectations could be interpreted as pessimism during specific episodes, such as when nominal interest rates are at the zero-lower-bound (Masolo and Monti (2015)). We use "pessimistic" as a short-hand labeling device for upwardly biased inflation expectations, regarding the interpretation as roughly right for households in most time-periods.

[13]Since the PCs and their factor loadings Λ are not separately identifiable, 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. This implies that the units for these series have no straight-forward interpretation in terms of the raw data.

Figure 5: Common and Heterogeneous Distortions



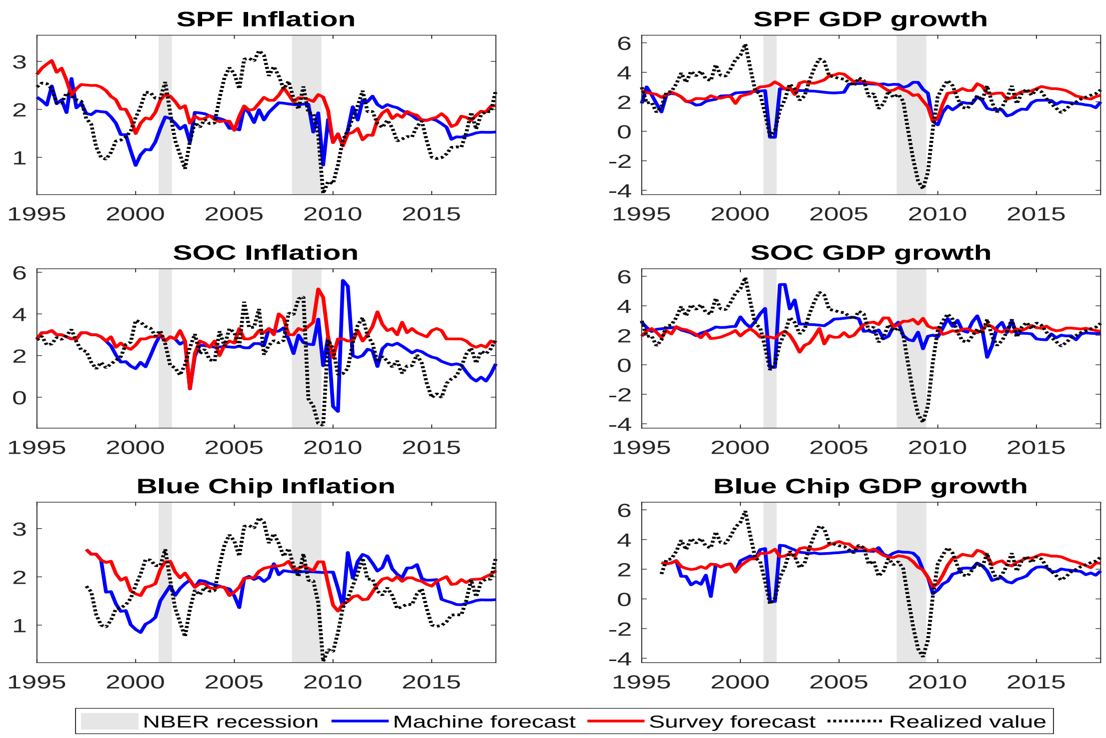
Common and heterogeneous belief distortions. The first row reports the first principal component of the biases across different surveys. For each respondent type, the second row reports the time series $bias^{(i)}_{j,t+h} = \mathbb{F}^{(i)}_t [y_{j,t+h}] - \mathbb{E}^{(i)}_t [y_{j,t+h}]$. The figure does not report the SOC GDP bias because only one series is available in that case. NBER recessions are shown with grey shaded bars. The sample is 1995:Q1-2018:Q2.

common component, as is a downward bias to inflation expectations for this same time period. At the same time, there is substantial heterogeneity across responses that varies over time, with greater dispersion observed in recessions. For the SPF, the most optimistic and pessimistic responses differ in some recession periods by more than 4% for GDP growth and by more than 2% for inflation, similarly for the BC survey. The high degree of disagreement among professional forecasters resulting in substantial heterogeneity in biases is an example of what Kahneman, Sibony, and Sunstein (2021) call "noise." For households in the SOC, the heterogeneity in measured belief distortions about inflation is enormous, especially immediately after the Great Recession, where the forecast of annual inflation from the respondent-type at the 95th percentile is almost 15%, while that for the respondent-type at the 5th percentile is less than -5%.

Figure 6 compares forecasted and actual values over time. The figure displays the median forecast of four-quarter-ahead inflation or GDP growth over our evaluation sample along with the actual inflation or GDP growth rate during the corresponding four quarter period being forecast. For all surveys, the machine has been more accurate not just on average over the long evaluation samples, but also consistently over the last five years of these samples, from 2013:Q2 to 2018:Q2, and by even larger magnitudes. For GDP growth, the ratio $MSE_{\mathbb{E}}/MSE_{\mathbb{F}}$ for the median SPF forecast is 0.70 over this subperiod, while it is 0.69 for median BC forecast. For inflation, the ratio $MSE_{\mathbb{E}}/MSE_{\mathbb{F}}$ over this same subperiod is 0.58 for the median SPF forecast, 0.67 for the median BC forecast, and 0.47 for the median SOC forecast. In both cases, professional forecasters and households alike under-performed in this subperiod because they systematically over-predicted both economic growth and inflation. That the machine does better at the end rather than the beginning of the evaluation sample is noteworthy, since it suggests that bounded rationality in the form of limitations on the human capacity for collecting and processing large amounts of information are unlikely to fully explain these findings. At least professional forecasters would have been capable by 2013 of taking advantage of advances in information-processing technology.

Unsurprisingly, the machine does not perform well in every time period of our sample. Figure 6 shows that professional forecasters made large forecast errors that were overly optimistic about GDP growth at the onset of the Great Recession, as noted in Gennaioli and Shleifer (2018). This pattern is likewise evident in Figure 6 for all surveys studied here. The figure shows that large forecast errors were made during this episode by the machine as well, with the machine algorithm 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. This occurs despite the fact that the machine algorithm takes into account hundreds of pieces of real-time information including that encoded in numerous financial series and dozens of credit spreads, recorded at daily, monthly, and quarterly sampling intervals. Such large *ex post* forecast errors during the Great Recession are arguably understandable when placed in the broader context

Figure 6: 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 machine forecast, and the realized inflation or GDP growth values during this period. Realized values are measured in real-time data as the vintage available four quarters after the period being forecast. NBER recessions are shown with grey shaded bars. The sample is 1995:Q1-2018:Q2.

of the time. As noted above, in contrast to many past recessions, yield spreads generally failed to signal the Great Recession to come. Moreover, the period leading up to the recession was characterized by unusually elevated objective uncertainty about the macroeconomy (Jurado, Ludvigson, and Ng (2015) and Ludvigson, Ma, and Ng (2019)). We argue that this episode underscores the role of largely unforeseen events in generating occasionally large prediction errors, not all of which can be attributed to a systematic bias in expectations.

Of course, with hindsight we now know that the Great Recession was preceded by a global financial crisis, itself triggered by a collapse in the value of residential real estate. It is thus tempting to consider feeding the machine a different switching indicator just prior to the Great Recession, for example one based on credit spreads or indicators of balance sheet health for firms and households. Our view is that this approach would be hard to defend, however. Unlike the case for yield spreads, where by the mid-1990s there existed a large body of public evidence showing their unique predictive power for recessions, there is no analogous body of empirical evidence for credit spreads and/or balance sheet indicators *before* 2007. Indeed, several of the empirical studies cited above and published in the early to mid 1990s explicitly compared credit spreads to yield spreads for forecasting output growth. Such studies universally found that credit spreads were comparatively weak predictors of recessions. Moreover, many of the balance sheet indicators now understood to be predictive for the global financial crisis (e.g., Greenwood, Hanson, Shleifer, and Sørensen (2020)) would not have been available in real time prior to the crisis, given the substantial data collection and data processing lags for such indicators. In short, the focus today on credit spread and balance sheet indicators as key predictors of recessions appears largely motivated by our *ex post* understanding of the 2007-2008 global financial crisis, rather than by a large body of prior knowledge that these indicators were the most robust predictors of economic contractions before the crisis. Since our approach is explicitly designed to exclude from the measure of belief distortions *ex post* mistakes that could only be understood with hindsight, we take the conservative approach of restricting our recession indicators to those that can be clearly defended on the basis of a prior body of publicly available empirical evidence.

5.3 Bias Decomposition

If the machine algorithm generates better forecasts, the survey respondents must be mis-weighting pertinent economic information. This raises the question: what kind of errors in judgement are the respondent-types making? To address this question, recall that the time t bias is defined as the difference between the survey respondent-type and machine forecasts:

$$\begin{aligned} bias_{j,t+h}^{(i)} &\equiv \mathbb{F}_{j,t+h|t}^{(i)} - \mathbb{E}_{j,t+h|t}^{(i)} = \mathbb{F}_t^{(i)}\left[y_{j,t+h}\right] - \hat{\alpha}_j - \hat{\beta}_{j\mathbb{F}}^{(i)}\mathbb{F}_t^{(i)}\left[y_{j,t+h}\right] - \hat{\mathbf{B}}_{j\mathcal{Z}}^{(i)\prime}\mathcal{Z}_{jt} \\ &= \underbrace{\left[-\hat{\alpha}_j^{(i)}\right]}_{\text{Intercept}} + \underbrace{\left[\left(1-\hat{\beta}_{j\mathbb{F}}^{(i)}\right)\mathbb{F}_t^{(i)}\left[y_{j,t+h}\right]\right]}_{\text{Survey}} + \underbrace{\left[-\hat{\mathbf{B}}_{j\mathcal{Z}}^{(i)\prime}\mathcal{Z}_{jt}\right]}_{\text{Info variables}} \end{aligned} \tag{9}$$

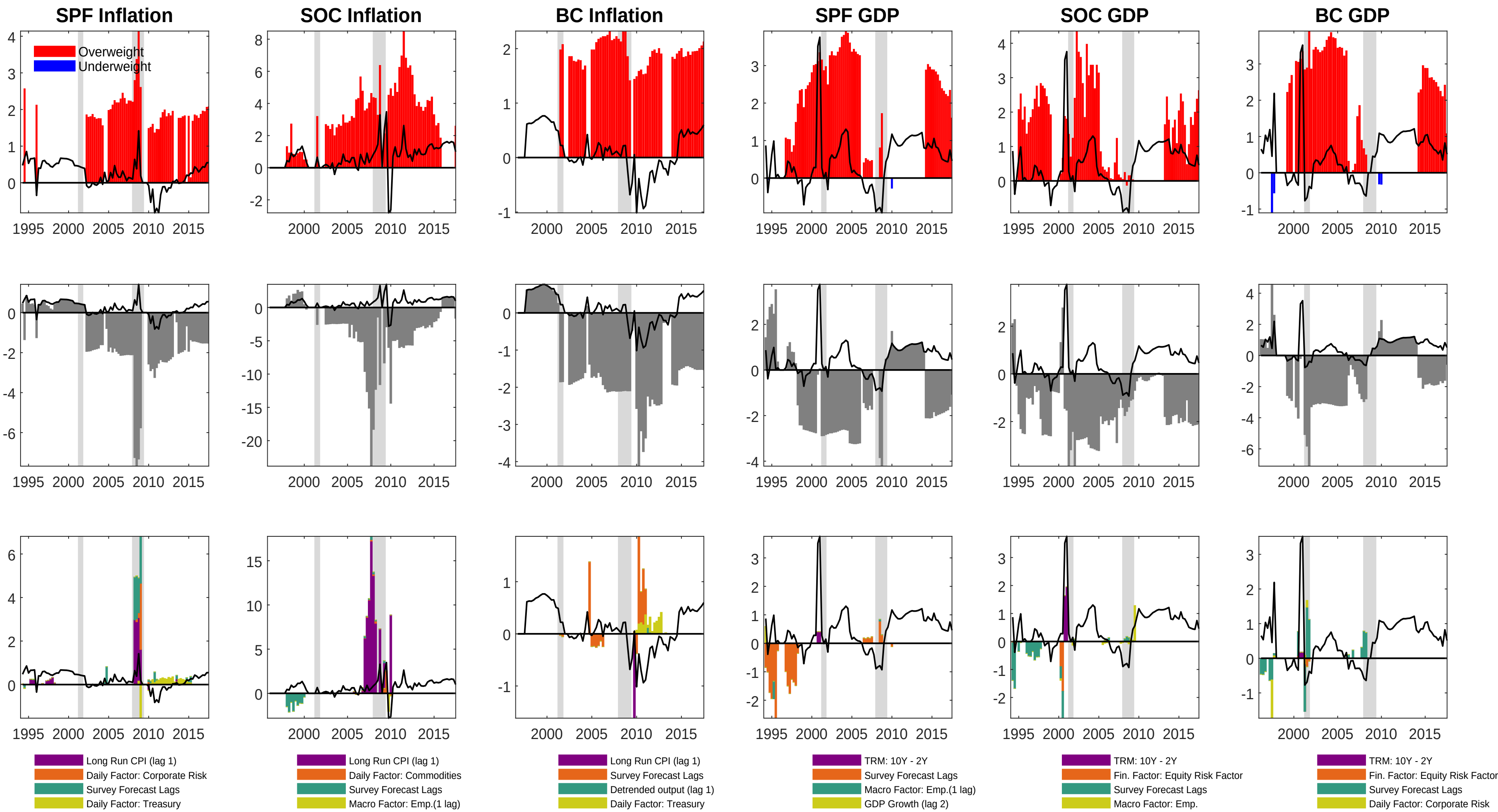
We are interested in the contribution of the three terms on the right-hand-side of (9), shown in large square brackets, the sum of which equals 100% of $bias_{j,t+h}^{(i)}$. This decomposition gives an indication of which information is most mis-weighted by the survey respondent-type, and by how much. The intercept term $\hat{\alpha}_j^{(i)}$ changes over the evaluation sample through the dynamic estimation algorithm and is akin to a time-varying latent conditional mean applied to the most recent rolling subsample window. We refer to this parameter as a "rolling mean" and denote it with a t subscript, i.e., $\hat{\alpha}_{j,t.}^{(i)}$ 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 likewise denoted with a t subscript.

It is useful to consider the magnitude and signs of the coefficients in the components above. First consider the coefficient on the survey forecast. If $\hat{\beta}_{j\mathbb{F},t}^{(i)} < 1$, this implies that the machine improves forecasts by downweighting the survey response in favor of giving greater absolute weight to publicly available information. Thus an estimate of $\hat{\beta}_{j\mathbb{F},t}^{(i)} < 1$ implies that the respondent-type *over*-weighted the marginal information in her own forecast relative to an efficient weighting of publicly available information. Conversely, if $\hat{\beta}_{j\mathbb{F},t}^{(i)} > 1$, the machine improved forecasts by giving greater weight to the survey forecast than the implicit weight given by the respondent-type to her own forecast. For the information variables and the rolling mean, any estimate of $\hat{\mathbf{B}}_{j\mathcal{Z},t}^{(i)\prime} \neq 0$ or $\hat{\alpha}_{j,t}^{(i)} \neq 0$ 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-type's implicit weight of zero conditional on her own forecast. Thus we refer to any estimate with $\hat{\mathbf{B}}_{j\mathcal{Z},t}^{(i)\prime} \neq 0$ or $\hat{\alpha}_{j,t}^{(i)} \neq 0$ as *under*-weighting of these sources of information.

Figure 7 reports, for each survey and each variable, the contribution to the bias in the median forecast of the three terms in square brackets in (9) at each point in time over our forecast evaluation samples. The solid lines in each subfigure of Figure 7 report the total median bias, $bias_{j,t+h}^{(50)}$, while the contributions of the three terms in square brackets in (9) are reported as bar charts, with the height of the bar showing the absolute magnitude by which that component contributed to the bias. Any above (below) zero bar indicates that the term contributed positively (negatively) to the overall bias. Since there are many terms in the information variable term, the figure reports contributions only for the most quantitatively important information variable contributors to the bias at each time t. In the case of the survey contribution, we further indicate with color coded bars whether a contribution to the bias was created by the respondent-type having over- or under-weighted her own forecast. A red bar indicates that the median respondent-type *over*-weighted her own forecast (i.e., $\hat{\beta}_{j\mathbb{F},t}^{(i)} < 1$), while a blue bar indicates that she *under*-weighted. For the intercept and information variables terms, any bar with a non-zero height indicates that the respondent-type gave too little absolute weight to that information. Recessions are shown in the figure by light grey shaded areas.

A key finding exhibited in Figure 7 that is robust across all surveys and all variables is that $\hat{\beta}_{j\mathbb{F},t}^{(50)}$ is very often substantially less than one. This happens not only for all surveys and for both inflation and GDP growth expectations, but also for most time periods in the evaluation sample. The mean (across time) values of $\hat{\beta}_{j\mathbb{F},t}^{(50)}$ for inflation and GDP growth are 0.40 and 0.41, respectively, for SPF, 0.33 and 0.52 for BC, and 0.14 and 0.41 for SOC. Thinking back to the model of public and private signals, we can interpret a finding of $\hat{\beta}_{j\mathbb{F},t}^{(i)} > 0$ as indicating that the marginal information contained in the survey response, capturing information intangible to the machine, such as that provided by a private signal or judgement, is in fact often valuable. But the finding that $\hat{\beta}_{j\mathbb{F},t}^{(i)} < 1$ indicates that such information is less valuable than the implicit weight placed on it by the survey respondent. The model of private and public signals implies that estimates $\hat{\beta}_{j\mathbb{F},t}^{(i)} < 1$ occur when the respondent-type over-relies on her private information or judgement, either because she over-estimates the precision of her private signal or because she inefficiently combines the public information, thereby under-estimating the precision of the public signal. Many instances of such an apparent overreliance are exhibited in Figure 7 by the frequent, tall red bars in the Survey Forecast panels of the first row. The length of the bars indicates that this factor contributes in most cases to quantitatively large distortions in macro expectations. For example, the first panel in Figure 7 indicates that this factor contributed 4% to the upward bias in the median SPF forecast of inflation–accounting for more than 100% of

Figure 7: Bias Decomposition: 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 bias, $F_t^{(50)}[y_{j,t+h}] - E_t^{(50)}[y_{j,t+h}]$. The barcharts in the first row panel report $\left(1 - \hat{\beta}_{j\mathbb{F}}^{(50)}\right)\mathbb{F}_t^{(50)}[y_{j,t+h}]$; those in the second row report $-\hat{\alpha}_j^{(50)}$; those in the third row report $-\hat{B}_{j\mathcal{Z}}^{(50)\prime}\mathcal{Z}_{jt}$ for the most important predictor contributors to the time t bias. 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 bias–during several periods at the end of the Great Recession.

If the median forecaster typically placed too much weight on her own forecast, then by definition she placed too little absolute weight on other information. The bottom two rows of Figure 7 gives an indication of the type of other objective economic information that was misweighted by the median forecaster over time. A key finding here is that the type of information is not static but instead changes over time. For example, in forming inflation expectations, the third rows shows that, during the Great Recession, too little attention was paid by the median SPF respondent to daily data on corporate credit spreads and to monthly data on long-run survey inflation forecasts, while for the years immediately after the Great Recession, between 2010 and 2015, the median respondent paid too little attention to daily information on Treasury yields and lagged values of the SPF forecasts. The type of information that was under-weighted varies also across surveys. For the SOC, under-weighting of long-run CPI survey forecasts shows up right before the Great Recession, but not elsewhere in the sample, while we find that the BC median forecast under-weighted this information *after* the Great Recession while subsequently giving too little weight to lagged survey forecasts.

Turning to expectations of economic growth, Figure 7 shows that the over-optimism displayed by professional forecasters (both SPF and BC) in the post-Great Recession period was largely driven, at first, by paying too little attention to the predictable slowing of average economic growth captured by the rolling mean, and then subsequently by repeated instances of over-weighting the marginal information in the survey response relative to what would be optimal under an efficient weighting of public information. Evidently, the median professional forecaster placed too much weight on a mistaken belief that economic growth would accelerate more than it did, a factor that accounts for more than 100% of the bias in the last five years of the sample. Overall the results are suggestive of a substantial overreliance by professional forecasters on the private or judgemental component of their predictions.

Taken together, the findings in Figure 7 underscore the crucial role of considering extensive and varied information in reducing forecaster bias. Although our machine learning algorithm often chooses sparse specifications, the findings in this figure show that *different* sparse information sets are relevant at different points in time. Since it is virtually impossible for a human to know with certainty which information may be relevant *ex ante*, algorithmic "openness" to wide-ranging and rich sources of information are vital for improving forecast accuracy over extended periods of time.

5.4 Some Comparisons With the Literature

With these results in hand, we now revisit some results in the prior literature that help illuminate the role played by key elements of our machine learning approach for establishing whether and by how much beliefs embedded in human judgements are distorted.

One key element pertains to the basic principle of out-of-sample versus in-sample forecasting,

Table 2: **CG Regressions of Forecast Errors on Forecast Revisions**

Panel A: **In-sample Regressions (CG Sample)**		
Regression: $\pi_{t+3} - \mathbb{F}_t^{(\mu)}[\pi_{t+3}] = \alpha^{(\mu)} + \beta^{(\mu)}\left(\mathbb{F}_t^{(\mu)}[\pi_{t+3}] - \mathbb{F}_{t-1}^{(\mu)}[\pi_{t+3}]\right) + \delta\pi_{t-1,t-2} + \epsilon_t$		
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)
$\pi_{t-1,t-2}$		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%.

a principle illustrated by contrasting results from *ex ante* and *ex post* econometric analyses, bearing in mind that survey respondents are asked to make genuine out-of-sample forecasts based on information known in real time. To illustrate the potential importance of this for the measurement of belief distortions, we revisit the *in-sample* regressions run in Coibion and Gorodnichenko (2015) (CG). CG found that mean survey forecast errors are positively predicted by *ex ante* mean forecast revisions in in-sample regressions. We reproduce their findings for the SPF on updated data and report the results in panel A of Table 2. Consistent with CG, we find strong evidence that lagged forecast revisions predict next period's forecast error in these regressions. Moreover, other information, e.g., lagged inflation, is found to be unimportant in predicting forecast errors once the information in forecast revisions is taken into account.[14] CG observe that these findings are consistent with the implications of theories that feature information frictions and under-reaction to aggregate news.[15]

The bottom panel of Table 2 reports results from the same regression forecasts, but this time run out-of-sample rather than in-sample. Over a range of forecast evaluation subsamples using either rolling or recursive regressions, we find that the mean SPF survey forecast generates much lower prediction error than a specification that attempts to exploit information in the lagged revision of the mean forecast. In other words, in contrast to the in-sample findings, information on lagged forecast revisions substantially *worsens* predictions of mean survey forecast errors in an out-of-sample context. This result recalls a body of prior econometric evidence finding that survey forecasts of inflation are hard to beat or even match with statistical models when forecasts are conducted out-of-sample.[16]

How can we reconcile the contradictory in-sample and out-of-sample evidence? One possibility is that the empirical relationship between forecast errors and lagged forecast revisions is highly unstable, as suggested by results below. Such an instability can create a high degree of sampling error so that what is revealed to be important *ex post* is simply not apparent *ex ante*. Whatever the reason for the poor performance of the specification out-of-sample, we've argued here that it is impossible to establish the extent to which beliefs are distorted due to information frictions or any other cause unless the benchmark against which distortions are measured adheres to the same forecasting context that survey respondents were faced with at the time they made their predictions. After all, even agents such as our machine who possess vast information processing capacity will optimally downweight information that might appear relevant *ex post* if it systematically failed to improve forecasts *ex ante*. It would not be correct to interpret this type of downweighting as under-reaction to economic news or as evidence of a systematic bias in expectations.

Next we ask whether lagged forecast revisions contain any valuable predictive information for forecast errors in our machine specifications. To do so, we run the following machine version of the CG regressions, which use the mean SPF forecast $\mathbb{F}^{(\mu)}$ and again placing observations on forecast errors on the left-hand-side:

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

This machine estimation differs from the CG estimation in three ways. First, the machine forecasts are made out-of-sample rather than in-sample. Second, the machine entertains the large-scale information set $\mathcal{Z}_{\pi t}$ as additional predictor variables. Third, the machine uses the EN estimator and dynamic cross-validation algorithm described above, while CG use least squares.

[14]We include one lag of the *quarterly* inflation rate as an additional control variable, consistent with the procedure implemented in CG. There is a typo in the published version of CG that erroneously indicates their procedure controlled for one lag of annual rather than quarterly inflation.

[15]As an aside, we note that the machine forecast errors do not exhibit a correlation with lagged machine forecast revisions, even in in-sample regressions. These results are reported in the Online Appendix.

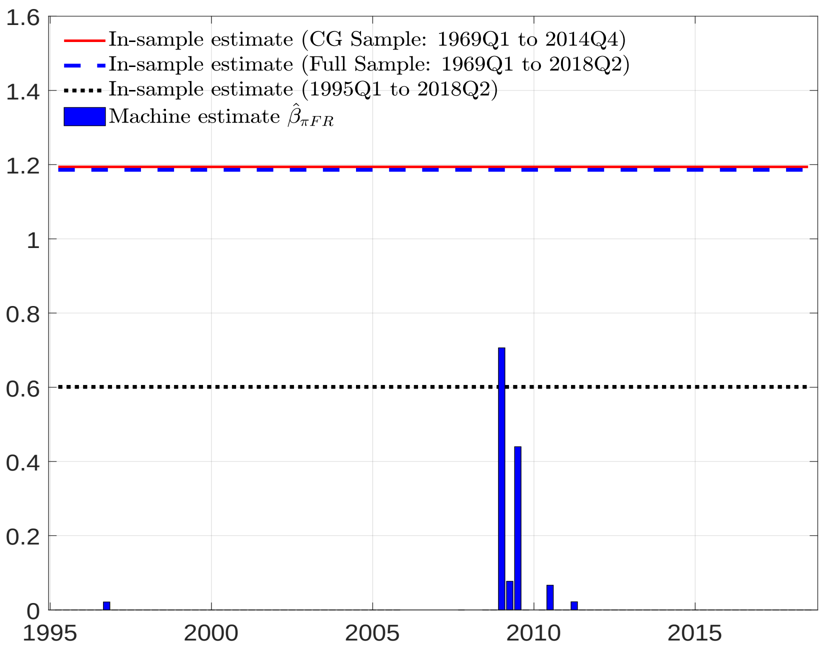
[16]For example, Ang, Bekaert, and Wei (2007), Del Negro and Eusepi (2011), Andersen, Bollerslev, Christoffersen, and Diebold (2011), Genre, Kenny, Meyler, and Timmermann (2013), and Faust and Wright (2013).

We denote the estimate of the coefficient on forecast revisions from this machine specification with $\beta_{\pi\text{FR}}^{(\mu)}$ and that from the univariate, in-sample least squares regression of CG as $\beta_{\pi\text{CG}}^{(\mu)}$.

Figure 8 reports the coefficients $\beta_{j\text{FR}}^{(\mu)}$ obtained from estimating (10) using the machine algorithm. 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 $\widehat{\beta}_{j\text{FR},t}^{(\mu)}$ using a bar chart, where the height of the bar indicates the magnitude of $\widehat{\beta}_{j\text{FR},t}^{(\mu)}$ and the time period t of the external evaluation sample 1995:Q1-2018:Q2 is given on the x-axis. Time periods τ for which there is no bar displayed indicate $\widehat{\beta}_{j\text{FR},\tau}^{(\mu)} = 0$. For comparison, in-sample estimates $\widehat{\beta}_{\pi\text{CG}}^{(\mu)}$ 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 external evaluation sample). The horizontal lines for $\widehat{\beta}_{\pi\text{CG}}^{(\mu)}$ over the first two samples are both close to 1.2, while that for the shorter recent sample are smaller by half. By contrast, the machine estimates $\widehat{\beta}_{j\text{FR},t}^{(\mu)}$ are always much smaller than the in-sample least squares estimates $\widehat{\beta}_{\pi\text{CG}}^{(\mu)}$ when those are obtained using the two longer subsamples, and they only match or exceed the half-as-large value in the shorter recent sample in one time period. Instead, the machine estimates of $\widehat{\beta}_{j\text{FR},t}^{(\mu)}$ are shrunk all the way to zero in 88 out of 94 quarters in favor of placing greater absolute weight on other pieces of information contained in $\mathcal{Z}_{\pi t}$ or $\widehat{\alpha}_{\pi,t}^{(\mu)}$. These findings do not indicate an important role for *ex ante* revisions in predicting average *ex post* forecast errors.

A second key element of our machine learning problem pertains to the data rich environment that survey respondents operate in. To illustrate the importance of this, we revisit an exercise in the spirit of Chauvet and Potter (2013), who considered a wide range of low dimensional statistical models for predicting GDP growth, finding that a second-order autoregression performed best for one-quarter ahead predictions when evaluated in a hold-out sample. Table 3 shows the estimated autoregressive coefficients from rolling, one-quarter-ahead, out-of-sample forecasting regressions of GDP growth on predictors, in two specifications. A high dimensional specification entertains a very large numbers of potential predictor variables, as in baseline machine specification. The two autoregressive lags are always among these predictors. A low dimensional specification uses the two autoregressive lags and only two additional predictors: the SPF median forecast of GDP growth and its current nowcast, both of which are also included in the high dimensional model. We find that the coefficient on the first autoregressive lag, large and positive in the low-dimensional setting, is zero in the high dimensional setting. As we have seen, this result does not imply that sparse specifications are rarely optimal. What it points to is the difficulty with knowing *which* small number of predictor variables are likely to be informative over time, when one does not have the benefit of hindsight afforded by an academic study of a single hold-out sample. The challenge for real-time decision making is that different pieces of information become relevant at different points in time.

Figure 8: Coefficient on Forecast Revisions



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

5.5 Belief Distortions over the Business Cycle

For our last set of results, we investigate the implications of our estimates for over- and under-reaction by survey respondents, a subject intense interest in the behavioral economics literature. We do so in a dynamic context in the wake of cyclical shocks, using the approach of Angeletos, Huo, and Sastry (2020) (AHS). Specifically, AHS estimate the dynamic responses of inflation or real GDP growth, as well as survey forecasts of those variables, to two cyclical shocks identified in Angeletos, Collard, and Dellas (2018a).[17] The cyclical shocks are the "inflation-targeted shock," ε_t^{π}, and the "GDP-targeted shock," ε_t^{GDP}. By construction, these shocks account for most of the business cycle variation in inflation and GDP growth, respectively.[18] Due to limitations of space, we restrict our reported results to belief distortions in the SPF median forecasts of four-quarter-ahead inflation or GDP growth.

Figure 9 reports dynamic responses of the machine forecast, the median SPF survey fore-

[17]We are grateful to the authors for providing us their data on these shocks.

[18]These shocks are identified using a 10 variable macro VAR as the structural shock that maximizes the volatility of the outcome variable (i.e., inflation, GDP growth) at frequencies corresponding to cycles between 6 and 32 quarters.

Table 3: **Average coefficients on the first two AR lags**

	High Dimensional	Low Dimensional
β_1	0.0000	0.0076
β_2	-0.0025	-0.0044

Note: This table reports 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.

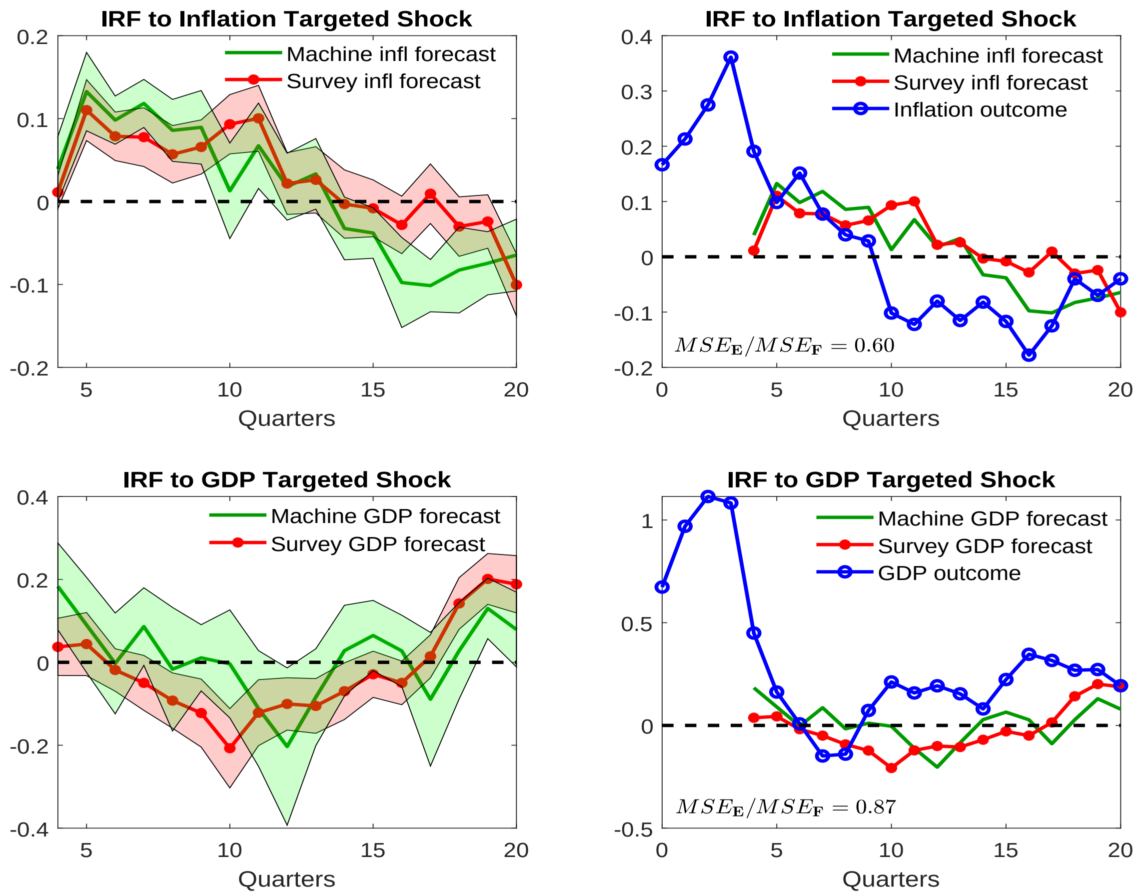
cast, and the relevant outcome variable, to innovations in ε_t^{π} and ε_t^{GDP}, estimated using local projections (Jorda (2005)).[19] The first column, first row, reports the responses of the machine and survey forecasts of inflation to an innovation in ε_t^{π}, while the first column, second row, reports the responses of the machine and survey forecasts of GDP growth to an innovation in ε_t^{GDP}. The right column shows these same responses along with the response of the relevant outcome variable, i.e., inflation or real GDP growth, removing the error bands to eliminate clutter. The plots in the right column "align" the forecast responses so that, at a given vertical slice of the plot, the outcome and forecast responses are measured over the same time horizon and the difference between the two is the forecast error. For example, given a shock at time t, the first response plotted for the survey forecast is $\mathbb{F}_t^{(50)}[y_{t+4}]$, which is aligned vertically with the response of y at time $t+4$. Following AHS, we set $H = 20$ quarters as the maximum period for tracing out impulse responses. Several findings from Figure 9 are worthy of emphasis.

First, survey respondents initially under-react to a shock but later over-react. Dynamic under- and over-reaction of the survey respondent's belief is measured vis-a-vis the machine belief–this is what is shown in the left column of the figure. From the left-hand subplots we observe that the survey forecast $\mathbb{F}_t^{(50)}[y_{t+4}]$ reacts less initially to an increase in both ε_t^{π} and ε_t^{GDP} than the machine forecast $\mathbb{E}_t^{(50)}[y_{t+4}]$ does, but eventually it reacts more. Qualitatively, these results are consistent with the dynamic patterns of initial under-reaction but delayed over-reaction emphasized by AHS.

Second, GDP growth expectations exhibit greater and more protracted *under*-reaction than do expectations about inflation. Conversely, inflation expectations exhibit greater and more protracted delayed *over*-reaction than do expectations about economic growth. In fact, for inflation expectations, the eventual over-reaction appears to be more important than the initial under-reaction, while GDP growth expectations appear to be more subject overall to under-reaction and only exhibit substantial over-reaction starting about 18 quarters after the shock.

[19] The Appendix gives the details of this estimation. We use a four-quarter forecast horizon, in contrast to AHS who use a three-quarter horizon. Our sample is also shorter than that used in AHS. The Appendix shows that we reproduce the results in AHS for the same forecast horizon and sample size that they use, and that the results are similar using the shorter sample of this paper.

Figure 9: Dynamic Responses to Cyclical Shocks



Dynamic responses of beliefs to cyclical shocks. The figure plots dynamic responses of the machine and survey beliefs $\mathbb{F}_t^{(50)}[\cdot]$ and $\mathbb{E}_t^{(50)}[\cdot]$ for the median respondent of the SPF to cyclical shocks from Angeletos, Collard, and Dellas (2018a) (AHS). The AHS inflation and GDP growth "targeted" cyclical shocks are those from a 10-variable VAR that maximize the volatility in inflation and GDP growth at business cycle frequencies, respectively. The right column aligns the forecast responses such that, at a given vertical slice, the outcome and forecast responses are measured over the same horizon, and the difference between the two is the forecast error. "$\text{MSE}_E/\text{MSE}_F$" is the ratio of the machine to survey mean-squared-forecast error averaged over the response time periods in the plot. The vintage of observations on the outcome variable is the one available four quarters after the period being forecast. Shaded areas are 68% confidence intervals based on HAC standard errors with a Bartlett kernel using four quarterly lags. The sample is 1995:Q1-2018:Q2.

Third, comparing the survey forecast to the *realized* value of the outcome variable greatly overstates the degree of over- or under-reaction that can be attributed to belief distortions. This can be observed in the right column of Figure 9 by noting that the first survey forecasts recorded after the shock under-shoot the realized outcome by much more than they under-shoot the machine forecasts. Likewise, the survey forecasts subsequently over-shoot the realized outcome by more than they over-shoot the machine forecast. AHS have interpreted the difference between the survey forecast and the realized value of the outcome variable as a measure of non-rational expectations. By contrast, we interpret the difference between the survey and *machine* forecasts as a measure of systematic expectational error, and the difference between the machine forecast and the outcome variable as pure random forecast error, rather than bias. The discrepancy between the two suggests that the cyclical shocks ε_t^{π} and ε_t^{GDP} are not well observed in real time, even by a machine with a high degree of information processing capacity. This may be because ε_t^{π} and ε_t^{GDP} are constructed from an in-sample estimation using fully revised, final release historical data, while both the survey and machine forecasts use only real-time information including that on the outcome variables being forecast, which are subject to significant processing and estimation delays.[20]

Fourth, in the wake of both cyclical shocks, the machine produces more accurate forecasts than the median SPF survey respondent. The right column of Figure 9 reports the ratio of the machine-to-survey MSE over the H periods for which the dynamic responses are tallied. The gains in forecast accuracy are especially large for inflation where the ratio $MSE_{\mathbb{E}}/MSE_{\mathbb{F}}$ is 0.6, but even for GDP growth the ratio $MSE_{\mathbb{E}}/MSE_{\mathbb{F}}$ is 0.87. That the machine improves forecasts in this context is noteworthy because it was not trained to optimize out-of-sample prediction at the specific business cycle frequencies that, by construction, dominate variation in the outcome variables in Figure 9.

6 Conclusion

This paper provides new measures of belief distortions in survey responses and relates them to macroeconomic activity. Our measures are based on a novel dynamic machine learning algorithm explicitly designed to combat overfitting and detect demonstrable, *ex ante* errors in macroeconomic expectations. For the median respondent from all surveys, expectations about both inflation and GDP growth are biased upward on average, with over-optimism about GDP growth especially prevalent among professional forecasters in the period after the Great Recession up to the end of our sample in 2018:Q2. 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 the marginal information embedded in their own belief and too little weight on other publicly available information. In response to cyclical shocks, we find that *under*-reaction preponderates in survey expectations of economic growth, while inflation expectations show greater delayed *over*-reaction. The results suggest that artificial intelligence algorithms can be productively deployed to correct errors in human judgement and improve predictive accuracy.

[20]The SPF collects survey responses in February on the outlook for GDP in the second quarter of the year, but the *advance* estimate of Q2 GDP is not released until the end of July. The *final release* data used to construct the shocks are subject to further subsequent revision over the course of two quarters. Some information pointing toward a large business cycle shock may be available at t, such as that contained in financial markets, but those are already accounted for by the machine.

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

Amisano, G., and J. Geweke (2017): "Prediction using several macroeconomic models," *Review of Economics and Statistics*, 99(5), 912–925.

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 (2018a): "Business cycle anatomy," Discussion paper, National Bureau of Economic Research.

——— (2018b): "Quantifying Confidence," *Econometrica*, 86(5), 1689–1726.

Angeletos, G.-M., Z. Huo, and K. A. Sastry (2020): "Imperfect Macroeconomic Expectations: Evidence and Theory," Discussion paper, National Bureau of Economic Research.

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.

Barber, B. M., and T. Odean (2000): "Trading is hazardous to your wealth: The common stock investment performance of individual investors," *The journal of Finance*, 55(2), 773–806.

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, Y. Ma, and A. Shleifer (2018): "Over-reaction in macroeconomic expectations," Discussion paper, National Bureau of Economic Research.

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.

Daniel, K., and D. Hirshleifer (2015): "Overconfident investors, predictable returns, and excessive trading," *Journal of Economic Perspectives*, 29(4), 61–88.

Daniel, K., D. Hirshleifer, and A. Subrahmanyam (1998): "Investor psychology and security market under-and overreactions," *the Journal of Finance*, 53(6), 1839–1885.

Daniel, K. D., D. Hirshleifer, and A. Subrahmanyam (2001): "Overconfidence, arbitrage, and equilibrium asset pricing," *The Journal of Finance*, 56(3), 921–965.

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.

Dotsey, M. (1998): "The predictive content of the interest rate term spread for future economic growth," *FRB Richmond Economic Quarterly*, 84(3), 31–51.

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

Estrella, A., and G. Hardouvelis (1990): "Possible roles of the yield curve in monetary analysis," *Intermediate Targets and Indicators for Monetary Policy.*

——— (1991): "The Term Structure as a Predictor of Real Economic Activity," *Journal of Finance*, 46(2), 555–76.

Estrella, A., and F. S. Mishkin (1998): "Predicting US recessions: Financial variables as leading indicators," *Review of Economics and Statistics*, 80(1), 45–61.

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.

Gabaix, X. (2014): "A sparsity-based model of bounded rationality," *The Quarterly Journal of Economics*, 129(4), 1661–1710.

——— (2020): "A behavioral New Keynesian model," *American Economic Review*, 110(8), 2271–2327.

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.

Giacomini, R., and H. White (2006): "Tests of conditional predictive ability," *Econometrica*, 74(6), 1545–1578.

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

Greenwood, R., S. G. Hanson, A. Shleifer, and J. A. Sørensen (2020): "Predictable financial crises," Discussion paper, National Bureau of Economic Research.

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

Hamilton, J. D. (1989): "A New Approach to the Analysis of Nonstationary Returns and the Business Cycle," *Econometrica*, 57, 357–384.

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.

Harvey, C. R. (1989): "Forecasts of Economic Growth from the Bond and Stock Markets," *Financial Analysts Journal*, 45(5), 38–45.

Haubrich, J. G., and A. M. Dombrosky (1996): "Predicting real growth using the yield curve," *Economic Review*, 32(1), 26–35.

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.

Jorda, O. (2005): "Estimation and Inference of Impulse Responses by Local Projections," *American Economic Review*, 95, 161–182.

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

Kahneman, D., O. Sibony, and C. Sunstein (2021): "Noise: a flaw in human judgment," .

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

Kozak, S., S. Nagel, and S. Santosh (2020): "Shrinking the cross-section," *Journal of Financial Economics*, 135(2), 271–292.

Kozicki, S., et al. (1997): "Predicting real growth and inflation with the yield spread," *Economic Review-Federal Reserve Bank of Kansas City*, 82, 39–58.

Ludvigson, S. C., S. Ma, and S. Ng (2019): "Uncertainty and Business Cycles: Exogenous Impulse or Endogenous Response?," *American Economic Journal: Macroeconomics,* forthcoming.

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.

——— (2010): "A Factor Analysis of Bond Risk Premia," in *Handbook of Empirical Economics and Finance*, ed. by A. Ulah, and D. E. A. Giles, vol. 1, pp. 313–72. Chapman and Hall, Boca Raton, FL.

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

——— (2016): "Learning from inflation experiences," *The Quarterly Journal of Economics*, 131(1), 53–87.

Mankiw, N. G., and R. Reis (2002): "Sticky information versus sticky prices: a proposal to replace the New Keynesian Phillips curve," *The Quarterly Journal of Economics*, 117(4), 1295–1328.

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

Masolo, R. M., and F. Monti (2015): "Ambiguity, Monetary Policy, and Trend Inflation," *Journal of the European Economic Association.*

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.

Morris, S., H. S. Shin, and M. Yildiz (2016): "Common belief foundations of global games," *Journal of Economic Theory*, 163, 826–848.

Newey, W. K., and K. D. West (1987): "A Simple, Positive Semidefinite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708.

Odean, T. (1998): "Volume, volatility, price, and profit when all traders are above average," *The journal of finance*, 53(6), 1887–1934.

Pesaran, M. H., and A. Timmermann (2007): "Selection of estimation window in the presence of breaks," *Journal of Econometrics*, 137(1), 134–161.

Plosser, C. I., and K. G. Rouwenhorst (1994): "International term structures and real economic growth," *Journal of monetary economics*, 33(1), 133–155.

Reis, R. (2006a): "Inattentive consumers," *Journal of monetary Economics*, 53(8), 1761–1800.

——— (2006b): "Inattentive producers," *The Review of Economic Studies*, 73(3), 793–821.

Sims, C. A. (2003): "Implications of rational inattention," *Journal of monetary Economics*, 50(3), 665–690.

Stock, J. H., and M. W. Watson (1989): "New Indexes of Coincident and Leading Economic Indicators," in *NBER Macroeconomics Annual: 1989*, ed. by O. J. Blanchard, and S. Fischer, pp. 351–94. MIT Press, Cambridge, MA.

——— (1991): "A Probability Model of the Coincident Economic Indicators," in *The Leading Economic Indicators: New Approaches and Forecasting Records*, ed. by G. Moore, and K. Lahiri, pp. 63–90. Cambridge University Press, Cambridge, MA.

——— (2002a): "Forecasting Using Principal Components From a Large Number of Predictors," *Journal of the American Statistical Association*, 97(460), 1167–79.

——— (2002b): "Macroeconomic Forecasting Using Diffusion Indexes," *Journal of Business and Economic Statistics*, 20(2), 147–62.

——— (2006): "Forecasting with Many Predictors," in *Handbook of Economic Forecasting, Volume 1*, ed. by M. H. Pesaran, and M. Weale, pp. 515–554. Elsevier, Oxford, United Kingdom.

Woodford, M. (2002): "Imperfect Common Knowledge and the Effects of Monetary Policy," in *In Knowledge, Information, and Expectations in Modern Macroeconomics: In Honor of Edmund S. Phelps*, ed. by P. Aghion, R. Frydman, J. Stiglitz, and M. Woodford, pp. 25–58. Princeton University Press, Cambridge MA.

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

Online Appendix

Additional Results

This section contains additional results not reported in the main text.

Table A.1: Machine Learning versus Survey Forecasts: Mean Forecasts

Machine Learning versus Survey Forecasts

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

Mean Forecast					
	Inflation			**GDP**	
	SPF	SOC	BC	SPF	BC
$MSE_{\mathbb{E}}/MSE_{\mu}$	0.95	0.42	0.84	0.89	0.76
OOS R^2	0.05	0.58	0.16	0.11	0.24

Machine v.s. survey mean-square-forecast errors. $MSE_{\mathbb{E}}$ and $MSE_{\mathbb{F}}$ denote the machine and survey mean-squared-forecast-errors, respectively, for 4-quarter-ahead forecasts, averaged over the evaluation sample. The out-of-sample Rsquared, OOS R^2, is defined as 1-$MSE_{\mathbb{E}}/MSE_{\mathbb{F}}$. The vintage of observations on the variable being forecast is the one available four quarters after the period being forecast. 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.

Table A.2: Machine Learning: K-Fold Estimation

Median SPF Inflation			
Center	Baseline	10-Fold , T_{IS}^*	10 Fold, Expanding Window
$MSE_{\mathbb{E}}/MSE_{\mathbb{F}}$	0.85	1.04	1.10

This table reports the MSE ratios between the machine forecast and survey median inflation forecast. The column "baseline" reports the ratio when machine forecasts are trained using optimal in-sample subsample of size T_{IS} and training and validation subsample of size T_{TS}. The second column reports the ratio when the machine forecasts are trained using 10-Fold cross validations over a rolling window of size T_{IS}. The third column reports the ratio when the machine forecasts are trained using 10-Fold cross validations over an expanding window.

Data

This appendix describes our data.

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.

Table A.3: Machine forecasts: Real-time vs. 2018:Q2 vintage

Machine vs Survey: Real-time vs 2018:Q2 Vintage

ML: $y_{j,t+h} = \alpha_j^{(\mu)} + \beta_{jF}^{(\mu)} F_t^{(\mu)} [y_{j,t+h}] + B_{jZ}^{(\mu)} Z_{jt} + \epsilon_{jt+h}$

SPF GDP Forecasts

Percentile	Median	Mean	5th	10th	20th	25th	30th
Real-time	0.88	0.91	0.70	0.81	0.80	0.84	0.87
Vintage 2018:Q2	0.79	0.83	0.68	0.76	0.75	0.77	0.80
Percentile	40th	60th	70th	75th	80th	90th	95th
Real-time	0.88	0.85	0.81	0.79	0.80	0.69	0.64
Vintage 2018:Q2	0.81	0.76	0.71	0.70	0.71	0.60	0.57

SPF Inflation Forecasts

Percentile	Median	Mean	5th	10th	20th	25th	30th
Real-time	0.85	0.95	0.56	0.74	0.83	0.90	0.88
Vintage 2018:Q2	0.79	0.89	0.56	0.69	0.75	0.82	0.79
Percentile	40th	60th	70th	75th	80th	90th	95th
Real-time	0.89	0.74	0.70	0.67	0.59	0.55	0.47
Vintage 2018:Q2	0.81	0.71	0.66	0.64	0.52	0.56	0.39

Machine v.s. survey mean-square-forecast errors. $MSE_{\mathbb{E}}$ and $MSE_{\mathbb{F}}$ denote the machine and survey mean-squared-forecast-errors, respectively, for 4-quarter-ahead forecasts, averaged over the evaluation sample. The vintage of observations on the variable being forecast is specified in the first column, either real-time as in the main text or using 2018Q2 vintage data. The evaluation period is 1995:Q1 to 2018:Q2.

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.

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.[21] 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.[22]

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

[21] 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.

[22] 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.

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."

   Since advance BEA estimates of these variables for the current quarter are unavailable at the time SPF respondents turn in their forecasts, four quarter-ahead inflation and GDP growth forecasts are constructed by dividing the forecasted level by the survey respondent-type's nowcast. Let $\mathbb{F}_t^{(i)}[P_{t+h}]$ be forecaster i's prediction of PGDP h quarters ahead and $\mathbb{N}_t^{(i)}[P_t]$ be forecaster i's nowcast of PGDP for the current quarter. Annualized inflation forecasts for forecaster i are

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

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

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

   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 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 out-of-sample 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.4 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.4: SPF Survey Deadlines[23]

Survey	Deadline Date	Survey	Deadline Date	Survey	Deadline Date
1990:Q1	Unknown	1991:Q1	2/16/91	1992:Q1	2/22/92
Q2	Unknown	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
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

[23]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.4 (Cont'd)

Survey	Deadline Date	Survey	Deadline Date	Survey	Deadline Date
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: To get a point forecast, we combine the information in the survey responses to questions A12 and A12b.

   - Question A12 asks (emphasis in original): *During the next 12 months, do you think that prices in general will go up, or go down, or stay where they are now?*
   - A12b asks (emphasis in original): *By about what percent do you expect prices to go (up/down) on the average, during the next 12 months?*

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 better, or worse 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. $\mathbb{F}_t^{(i)}\left[gP_{t+h}^{(Q/Q)}\right]$ are reported at annual rates in percentage points, so we convert to quarterly raw units before compounding. 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:

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

   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.[24] 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.

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

Following Coibion and Gorodnichenko (2015), to construct forecasts and forecast errors, we use the vintage of inflation and GDP growth data 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: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.

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 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 a measure of residential real estate prices from the Case-Shiller/S&P index deflated by the Consumer Price Index, and 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 the first month of that quarter's values. A few anomalous observations have unknown survey deadlines (e.g., the SPF deadlines for 1990:Q1). In such cases, we allow only information up to quarter $t-1$ to enter the model. Thus, we use the last month of the previous quarter's values in these cases.

Table A.5 gives the complete list of real-time macro variables. 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.[25] Table A.5 also lists the transformation applied to each variable to make them stationary before 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.5: 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

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

Table A.5 (Cont'd)

No.	Short Name	Source	Tran	Description	First Vintage
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, Investment, Housing		
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
38	CASESHILLER	S&P	Δln	Case-Shiller US National Home Price index/CPI	1987:M1
			Group 4: Consumption		
39	NCONGMMVMD	Philly Fed	Δln	Nom. personal cons. exp. - Goods	2009:M8
40	NCONHHMMVMD	Philly Fed	Δln	Nom. hh. cons. exp.	2009:M8
41	NCONSHHMMVMD	Philly Fed	Δln	Nom. hh. cons. exp. - Services	2009:M8
42	NCONSNPMMVMD	Philly Fed	Δln	Nom. final cons. exp. of NPISH	2009:M8
43	RCONDMMVMD	Philly Fed	Δln	Real personal cons. exp. - Durables	1998:M11
44	RCONGMMVMD	Philly Fed	Δln	Real personal cons. exp. - Goods	2009:M8
45	RCONHHMMVMD	Philly Fed	Δln	Real hh. cons. exp.	2009:M8
46	RCONMMVMD	Philly Fed	Δln	Real personal cons. exp. - Total	1998:M11
47	RCONNDMVMD	Philly Fed	Δln	Real personal cons. exp. - Nondurables	1998:M11
48	RCONSHHMMVMD	Philly Fed	Δln	Real hh. cons. exp. - Services	2009:M8
49	RCONSMMVMD	Philly Fed	Δln	Real personal cons. exp. - Services	1998:M11
50	RCONSNPMMVMD	Philly Fed	Δln	Real final cons. exp. of NPISH	2009:M8
51	NCONGMVQD	Philly Fed	Δln	Nom. personal cons. exp. - Goods	2009:Q3
52	NCONHHMVQD	Philly Fed	Δln	Nom. hh. cons. exp.	0209:Q3
53	NCONSHHMVQD	Philly Fed	Δln	Nom. hh. cons. exp. - Services	2009:Q3
54	NCONSNPMVQD	Philly Fed	Δln	Nom. final cons. exp. of NPISH	2009:Q3
55	RCONDMVQD	Philly Fed	Δln	Real personal cons. exp. - Durable goods	1965:Q4
56	RCONGMVQD	Philly Fed	Δln	Real personal cons. exp. - Goods	2009:Q3
57	RCONHHMVQD	Philly Fed	Δln	Real hh. cons. exp.	2009:Q3
58	RCONMVQD	Philly Fed	Δln	Real personal cons. exp. - Total	1965:Q4
59	RCONNDMVQD	Philly Fed	Δln	Real pesonal cons. exp. - Nondurable goods	1965:Q4
60	RCONSHHMVQD	Philly Fed	Δln	Real hh. cons. exp. - Services	2009:Q3
61	RCONSMVQD	Philly Fed	Δln	Real personal cons. exp. - Services	1965:Q4
62	RCONSNPMVQD	Philly Fed	Δln	Real final cons. exp. of NPISH	2009:Q3
63	NCONQVQD	Philly Fed	Δln	Nom. personal cons. exp.	1965:Q4
			Group 5: Prices		
64	PCONGMMVMD	Philly Fed	$\Delta^2 ln$	Price index for personal cons. exp. - Goods	2009:M8
65	PCONHHMMVMD	Philly Fed	$\Delta^2 ln$	Price index for hh. cons. exp.	2009:M8
66	PCONSHHMMVMD	Philly Fed	$\Delta^2 ln$	Price index for hh. cons. exp. - Services	2009:M8
67	PCONSNPMMVMD	Philly Fed	$\Delta^2 ln$	Price index for final cons. exp. of NPISH	2009:M8
68	PCPIMVMD	Philly Fed	$\Delta^2 ln$	Consumer price index	1998:M11
69	PCPIXMVMD	Philly Fed	$\Delta^2 ln$	Core consumer price index	1998:M11
70	PPPIMVMD	Philly Fed	$\Delta^2 ln$	Producer price index	1998:M11
71	PPPIXMVMD	Philly Fed	$\Delta^2 ln$	Core producer price index	1998:M11
72	PCONGMVQD	Philly Fed	$\Delta^2 ln$	Price index for personal. cons. exp. - Goods	2009:Q3

Table A.5 (Cont'd)

No.	Short Name	Source	Tran	Description	First Vintage
73	PCONHHMVQD	Philly Fed	$\Delta^2 ln$	Price index for hh. cons. exp.	2009:Q3
74	PCONSHHMVQD	Philly Fed	$\Delta^2 ln$	Price index for hh. cons. exp. - Services	2009:Q3
75	PCONSNPMVQD	Philly Fed	$\Delta^2 ln$	Price index for final cons. exp. of NPISH	2009:Q3
76	PCONXMVQD	Philly Fed	$\Delta^2 ln$	Core price index for personal cons. exp.	1996:Q1
77	CPIQVMD	Philly Fed	$\Delta^2 ln$	Consumer price index	1994:Q3
78	PQVQD	Philly Fed	$\Delta^2 ln$	Price index for GNP/GDP	1965:Q4
79	PCONQVQD	Philly Fed	$\Delta^2 ln$	Price index for personal cons. exp.	1965:Q4
80	PIMPQVQD	Philly Fed	$\Delta^2 ln$	Price index for imports of goods and services	1965:Q4
		Group 6: Trade and Government			
81	REXMVQD	Philly Fed	Δln	Real exports of goods and services	1965:Q4
82	RGMVQD	Philly Fed	Δln	Real government cons. and gross inv. - Total	1965:Q4
83	RGFMVQD	Philly Fed	Δln	Real government cons. and gross inv. - Federal	1965:Q4
84	RGSLMVQD	Philly Fed	Δln	Real government cons. and gross. inv. - State and local	1965:Q4
85	RIMPMVQD	Philly Fed	Δln	Real imports of goods and services	1965:Q4
86	RNXMVQD	Philly Fed	Δlv	Real net exports of goods and services	1965:Q4
		Group 7: Money and Credit			
87	BASEBASAQVMD	Philly Fed	$\Delta^2 ln$	Monetary base	1980:Q2
88	M1QVMD	Philly Fed	$\Delta^2 ln$	M1 money stock	1965:Q4
89	M2QVMD	Philly Fed	$\Delta^2 ln$	M2 money stock	1971:Q2
90	NBRBASAQVMD	Philly Fed	$\Delta lv/lv$	Nonborrowed reserves	1967:Q3
91	NBRECBASAQVMD	Philly Fed	$\Delta lv/lv$	Nonborrowed reserves plus extended credit	1984:Q2
92	TRBASAQVMD	Philly Fed	$\Delta^2 ln$	Total reserves	1967:Q3
93	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 $\sqrt{T}/N \to \infty$), 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.

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 the first month of that quarter's values. Alternatively, a few anomalous observations have unknown survey deadlines (e.g., the SPF deadlines for 1990:Q1). In such cases, we allow only information up to quarter $t-1$ to enter the model. Thus, we use the last month of the previous quarter's values in these cases.

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.6: List of Financial Dataset Variables

No.	Short Name	Source	Tran	Description
		Group 1: Prices, Yield, Dividends		

Table A.6 (Cont'd)

No.	Short Name	Source	Tran	Description
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)
63	2_2	Kenneth French	lv	(2, 2) portfolio sorted on (size, book-to-market)

Table A.6 (Cont'd)

No.	Short Name	Source	Tran	Description
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)
129	8_9	Kenneth French	*lv*	(8, 9) portfolio sorted on (size, book-to-market)

Table A.6 (Cont'd)

No.	Short Name	Source	Tran	Description
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_libra

- 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.

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.7 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.7: 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

Table A.7 (Cont'd)

No.	Short Name	Source	Tran	Description
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
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

Table A.7 (Cont'd)

No.	Short Name	Source	Tran	Description
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
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, $\mathbf{G}^Q_{D,t}$. Denote by G^D_t a factor in a daily frequency formed from the daily financial dataset and denote by G^Q_t a quarterly aggregate of the corresponding daily factor time series. Let $G^D_{N_D-j,d_t,t}$ 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 $G^D_{N_D,d_t,t}$. For simplicity, we suppress the subscript d_t thus $G^D_{N_D-j,d_t,t} \equiv G^D_{N_D-j,t}$.

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^Q_t

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:

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

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 $w(i; \theta_1, \theta_2)$ 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 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 $\boldsymbol{\lambda}_t^{(i)} = \left(\lambda_{1t}^{(i)}, \lambda_{2t}^{(i)}\right)'$ (See the next section for a description of EN). The procedure involves iterating on the steps given in the main text.

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

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

If the machine model is implemented as an estimation with using forecast errors as the dependent variable, i.e.,

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

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 EN shrinks estimated coefficients toward zero, this results in shrinkage of $\beta_{j\mathbb{F}}^{(i)}$ toward unity. In this case the machine forecast is given by

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

By contrast, if the machine forecast is implemented by running the specification

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

then $\beta_{j\mathbb{F}}^{(i)}$ is shrunk toward zero and the algorithm will typically place less weight on the survey forecast than the specification (A.12). 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 $\mathcal{L}\left(\boldsymbol{\lambda}_t^{(i)}, T_{IS}, T_{TS}\right)$ over psuedo out-of-sample forecast errors in every training sample.

Elastic Net Estimator

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 $\boldsymbol{\beta}_j = \left(\alpha_j, \beta_{j\mathbb{F}}, \text{vec}\left(\mathbf{B}_{j\mathcal{Z}}\right)\right)' \equiv (\beta_0, \beta_1, ...\beta_K)'$ 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:

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

By minimizing the MSE over the training samples, we choose the optimal $\lambda_1^{(i)}$ and $\lambda_2^{(i)}$ values simultaneously.

Dynamic Factor Estimation

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"), $\boldsymbol{\Lambda}_i^C$ is a corresponding $r_C \times 1$ vector of latent factor loadings, and e_{it}^X is a vector of idiosyncratic errors.[26] 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 to account for nonlinearities are formed by including polynomial functions of $\mathbf{G}_t^C$, and by including factors formed from polynomials of the raw data.

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

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.[27] 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 $\mathbb{E}(\widetilde{x}_{it}) = \mathbb{E}(\lambda_i' \hat{g}_t)$, 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 $\mathbf{G}^2_{raw1}$.

[27] 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:

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

where M is a constant.

7. Our matrix of factors is $[\mathbf{G}_{raw}, \mathbf{G}_{sqr1}, \mathbf{G}_{raw1}^2]$, 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.5. 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.[28]

For the financial factors, we use all of the variables listed in Table A.6, and no additional steps are performed beyond those described above.

Economic Names of 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.8 reports the series with largest average marginal R^2 for each factor of each large dataset.

Predictor Variables

The vector $\mathbf{Z}_{jt} \equiv \left(y_{j,t}, \hat{\mathbf{G}}_t', \mathbf{W}_{jt}'\right)'$ is an $r = 1 + r_G + r_W$ vector which collects the data at time t with $\mathcal{Z}_{jt} \equiv \left(y_{j,t}, ..., y_{j,t-p_y}, \hat{\mathbf{G}}_t', ..., \hat{\mathbf{G}}_{t-p_G}', \mathbf{W}_{jt}', ..., \mathbf{W}_{jt-p_W}'\right)'$ 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. Superscript (i) refers 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 $\mathbf{W}_{jt}'$:

[28]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.

Table A.8: 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: Housing
$G_{7,M,t}$	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.

1. $\mathbb{F}^{(i)}_{jt-k}\left[y_{jt+h-k}\right]$, lagged values of the ith type's forecast, where $k = 1, 2, ...$

2. $\mathbb{F}^{(s \neq i)}_{jt-1}\left[y_{jt+h-1}\right]$, lagged values other type's forecasts, $s \neq i$

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

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

5. Trend inflation measured as $\overline{\pi}_{t-1} = \begin{cases} \rho\overline{\pi}_{t-2} + (1-\rho)\pi_{t-1},\ \rho = 0.95 & \text{if } t < \text{1991:Q4} \\ \text{CPI10}_{t-1} & \text{if } t \geq \text{1991:Q4}, \end{cases}$ where CPI10 is the median SPF forecast of annualized average inflation over the current and next nine years. 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. $\mathbf{G}_{D,t}^Q$, 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.[29] 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.

[29] 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

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

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

1. $\mathbb{F}^{(i)}_{jt-k}\left[y_{jt+h-k}\right]$, lagged values of the ith type's forecast, where $k = 1, 2, ...$
2. $\mathbb{F}^{(s\neq i)}_{jt-1}\left[y_{jt+h-1}\right]$, lagged values other type's forecasts, $s \neq i$
3. $\text{var}_N\left(\mathbb{F}^{(\cdot)}_{t-1}\left[y_{jt+h-1}\right]\right)$, where $\text{var}_N\left(\cdot\right)$ denotes the cross-sectional variance of forecasts
4. $\text{skew}_N\left(\mathbb{F}^{(\cdot)}_{t-1}\left[y_{jt+h-1}\right]\right)$, where $\text{skew}_N\left(\cdot\right)$ denotes the cross-sectional skewness of forecasts
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. $\mathbb{N}^{(i)}_{t}\left[y_{t,t-h}\right]$ = 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. $\mathbf{G}^Q_{D,t}$, 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 $\mathbf{G}_{D,t}\left(\mathbf{w}\right)$ and the daily factors are aggregated up to quarterly observations $\mathbf{G}^Q_{D,t}\left(\mathbf{w}\right)$ 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 $\mathbf{G}_{D,t}\left(\mathbf{w}\right)$.

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.[30] 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 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,

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

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 $\mathbb{F}_{js}^{SPF,(i)}[\cdot]$ and from the Michigan Survey using $\mathbb{F}_{js}^{MS,(i)}[\cdot]$.

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

1. $\mathbb{F}_{jt-1}^{SPF,(\mu)}[y_{jt+h-1}]$, the mean SPF forecast for CPI.
2. $\mathbb{F}_{jt-1}^{SPF,(50)}[y_{jt+h-1}]$, the $50th$ percentile SPF forecast for CPI.
3. $\mathbb{F}_{jt-1}^{SPF,(25)}[y_{jt+h-1}]$, the $25th$ percentile SPF forecast for CPI.
4. $\mathbb{F}_{jt-1}^{SPF,(75)}[y_{jt+h-1}]$, the $75th$ percentile SPF forecast for CPI.
5. $\text{var}_N\left(\mathbb{F}_{t-1}^{SPF,(\cdot)}[y_{jt+h-1}]\right)$, the cross-sectional variance of SPF forecasts of CPI.
6. $\text{skew}_N\left(\mathbb{F}_{t-1}^{SPF,(\cdot)}[y_{jt+h-1}]\right)$, the cross-sectional skewness of SPF forecasts of CPI.
7. Trend inflation measured as $\overline{\pi}_{t-1} = \begin{cases} \rho\overline{\pi}_{t-2} + (1-\rho)\pi_{t-1}, \ \rho = 0.95 & \text{if } t < \text{1991:Q4} \\ \text{CPI10}_{t-1} & \text{if } t \geq \text{1991:Q4}, \end{cases}$ where CPI10 is the median SPF forecast of annualized average inflation over the current and next nine years. Trend inflation is intended to capture long-run trends. When long-run

[30] 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

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.[31] 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.

[31] 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 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).

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.[32] 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.

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.

Private and Public Signals

This section derives the coefficient estimates given in the text for a model in which forecasters are presumed to combine statistical predictive models using public information (i.e., public signals) with their own judgemental forecast (i.e., a private signal) to form an overall prediction.

For reference, the machine learning model is

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

To set the stage, suppose that we have a jointly normal random vector split into two pieces, X_1 and X_2:

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \right)$$

Denote the conditional distribution of X_1 given X_2 as $X_1|X_2$. With joint normality the optimal updating rule is

$$X_1|X_2 \sim N\left(\mu_1 + \beta\left(X_2 - \mu_2\right), \Omega\right)$$

[32]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

$$\beta = \Sigma_{12}\Sigma_{22}^{-1}, \qquad \Omega = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}.$$

Below we use $\mathbb{E}_o[\cdot]$ to denote the conditional expectation implied by the optimal updating rule, i.e., $\mathbb{E}_o[X_1|X_2] = \mu_1 + \beta(X_2 - \mu_2)$ in the above. These results are used below.

Let x be publicly available information and let z be a private signal about an unknown variable y. These variables are related to one another according to the system

$$\begin{aligned} x &\sim iid\left(0, \sigma_x^2\right) \\ y &= \alpha x + u_2,\ u_2 \sim N\left(0, \sigma_2^2\right) \qquad \text{(A.14)} \\ z &= y + u_1,\ u_1 \sim N\left(0, \sigma_1^2\right), \qquad \text{(A.15)} \end{aligned}$$

where α is a known parameter describing the mapping from x to y, and x, u_1, and u_2 are i.i.d. and mutually uncorrelated with one another.

Consider the optimal forecast of y in this setting, when one combines a statistical model using x with a private signal z. From (A.14) and (A.15), conditional on observing αx we have

$$\begin{aligned} y|x &= \alpha x + u_2,\ u_2 \sim N\left(0, \sigma_2^2\right) \\ z &= (y|x) + u_1 = \alpha x + u_2 + u_1,\ u_1 \sim N\left(0, \sigma_2^2\right) \end{aligned}$$

$$\begin{bmatrix} y|x \\ z|x \end{bmatrix} \sim N\left(\begin{bmatrix} \alpha x \\ \alpha x \end{bmatrix}, \begin{bmatrix} \sigma_2^2 & \sigma_2^2 \\ \sigma_2^2 & \sigma_1^2 + \sigma_2^2 \end{bmatrix}\right)$$

Now suppose that the agent combines the information in αx with a private signal z.[33] Conditional on both private and public signals, the optimal forecast is

$$\begin{aligned} \mathbb{E}_o[y|x,z] &= \alpha x + \sigma_2^2\left(\sigma_1^2 + \sigma_2^2\right)^{-1}(z - \alpha x) \\ &= \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} z + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2}\alpha x \\ &= \gamma z + (1 - \gamma)\alpha x \end{aligned}$$

where

$$\gamma \equiv \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}.$$

The optimal weight on the private signal z versus the public signal αx depends on the true precision σ_1^{-2} of the private signal relative the precision σ_2^{-2} of the public signal.

We can also compute:

$$\begin{aligned} \mathbb{V}_o[y|x,z] &= \sigma_2^2 - \sigma_2^2\sigma_2^2\left(\sigma_1^2 + \sigma_2^2\right)^{-1} \\ &= \sigma_2^2\left[1 - \sigma_2^2\left(\sigma_1^2 + \sigma_2^2\right)^{-1}\right] \\ &= \sigma_2^2\sigma_1^2\left(\sigma_1^2 + \sigma_2^2\right)^{-1}. \end{aligned}$$

[33] The result below is unchanged if one assumes that the agent first recieves z and then combines it with αx.

Forecaster. The forecaster assigns weights to the private and public signal as follows:

$$\mathbb{F} = \gamma^F z + \left(1 - \gamma^F\right) \alpha^F x$$

The survey response is here interpreted as a forecast based partly on a respondent's statistical model using public information ($\alpha^F z_2$) in combination with a private signal z.

Machine. The machine forecast of y given by

$$\mathbb{E} = \widehat{\beta}\mathbb{F} + \widehat{B}x = \widehat{\beta}\left[\gamma^F z + \left(1 - \gamma^F\right) \alpha^F x\right] + \widehat{B}x,$$

where $\widehat{\beta}$ and $\widehat{B}$ are estimated coefficients. The machine uses maximum likelihood to estimate the coefficients $b \equiv (\beta, B)'$ from a regression of y on $f \equiv (\mathbb{F}, x)'$, with $b =$cov$(f, f')^{-1}$cov(f, y). This estimator results in the values:

$$\begin{aligned} \widehat{\beta} &= \frac{\gamma}{\gamma^F} \\ \widehat{B} &= (1 - \gamma)\alpha - \left(\frac{\gamma}{\gamma^F} - \gamma\right)\alpha^F. \end{aligned}$$

Proof:

Writing the covariances in matrix form, we have that

$$b = \begin{bmatrix} \beta \\ B \end{bmatrix} = \begin{bmatrix} \text{var}(\mathbb{F}) & \text{cov}(\mathbb{F},\, x) \\ \text{cov}(\mathbb{F},\, x) & \text{var}(x) \end{bmatrix}^{-1} \begin{bmatrix} \text{cov}(\mathbb{F}, y) \\ \text{cov}(x, y) \end{bmatrix}. \tag{A.16}$$

First,

$$\begin{aligned} \text{var}(\mathbb{F}) &= \text{var}\left(\gamma^F z + \left(1 - \gamma^F\right)\alpha^F x\right) \\ &= \text{var}\left(\gamma^F(\alpha x + u_2 + u_1) + \left(1 - \gamma^F\right)\alpha^F x\right) \\ &= \left(\gamma^F\right)^2\left(\alpha^2\sigma_x^2 + \sigma_2^2 + \sigma_1^2\right) + \left[\left(1 - \gamma^F\right)\alpha^F\right]^2\sigma_x^2 \\ &= \left(\gamma^F\right)^2\left(\sigma_2^2 + \sigma_1^2\right) + \left(\gamma^F\alpha + \left(1 - \gamma^F\right)\alpha^F\right)^2\sigma_x^2 \end{aligned} \tag{A.17}$$

Second,

$$\begin{aligned} \text{cov}(\mathbb{F},\, x) &= \text{cov}\left(\left(\gamma^F\alpha + \left(1 - \gamma^F\right)\alpha^F\right)x + \gamma^F(u_2 + u_1), x\right) \\ &= \left(\gamma^F\alpha + \left(1 - \gamma^F\right)\alpha^F\right)\sigma_x^2 \end{aligned} \tag{A.18}$$

Third,

$$\text{var}(x) = \sigma_x^2 \tag{A.19}$$

Fourth,

$$\begin{aligned} \text{cov}(\mathbb{F}, y) &= \text{cov}\left(\left(\gamma^F\alpha + \left(1 - \gamma^F\right)\alpha^F\right)x + \gamma^F(u_2 + u_1), \alpha x + u_2\right) \\ &= \left(\gamma^F\alpha + \left(1 - \gamma^F\right)\alpha^F\right)\alpha\sigma_x^2 + \gamma^F\sigma_2^2 \end{aligned} \tag{A.20}$$

Last,

$$
\begin{aligned}
cov(x, y) &= cov(x, \alpha x + u_2) \\
&= \alpha \sigma_x^2. \qquad \text{(A.21)}
\end{aligned}
$$

Plugging equation (A.17) to (A.21) to equation (A.16), we have

$$
\begin{aligned}
b &= \underbrace{\begin{bmatrix} \left(\gamma^F\right)^2 (\sigma_2^2 + \sigma_1^2) + \left(\gamma^F \alpha + \left(1 - \gamma^F\right) \alpha^F\right)^2 \sigma_x^2 & \left(\gamma^F \alpha + \left(1 - \gamma^F\right) \alpha^F\right) \sigma_x^2 \\ \left(\gamma^F \alpha + \left(1 - \gamma^F\right) \alpha^F\right) \sigma_x^2 & \sigma_x^2 \end{bmatrix}^{-1}}_{\equiv B_1} \\
&\quad \times \begin{bmatrix} \left(\gamma^F \alpha + \left(1 - \gamma^F\right) \alpha^F\right) \alpha \sigma_x^2 + \gamma^F \sigma_2^2 \\ \alpha \sigma_x^2. \end{bmatrix} \\
&= \frac{1}{\Delta} \begin{bmatrix} \sigma_x^2 & -\left(\gamma^F \alpha + \left(1 - \gamma^F\right) \alpha^F\right) \sigma_x^2 \\ -\left(\gamma^F \alpha + \left(1 - \gamma^F\right) \alpha^F\right) \sigma_x^2 & \left(\gamma^F\right)^2 (\sigma_2^2 + \sigma_1^2) + \left(\gamma^F \alpha + \left(1 - \gamma^F\right) \alpha^F\right)^2 \sigma_x^2 \end{bmatrix} \\
&\quad \times \begin{bmatrix} \left(\gamma^F \alpha + \left(1 - \gamma^F\right) \alpha^F\right) \alpha \sigma_x^2 + \gamma^F \sigma_2^2 \\ \alpha \sigma_x^2. \end{bmatrix}
\end{aligned}
$$

where Δ is the determinant of the matrix B_1.

We then have,

$$
\begin{aligned}
\Delta \beta &= \left[\left(\gamma^F \alpha + \left(1 - \gamma^F\right) \alpha^F\right) \alpha \sigma_x^4 + \gamma^F \sigma_2^2 \sigma_x^2 - \alpha \left(\gamma^F \alpha + \left(1 - \gamma^F\right) \alpha^F\right) \sigma_x^2 \sigma_x^2\right] \\
&= \gamma^F \sigma_2^2 \sigma_x^2 \qquad \text{(A.22)}
\end{aligned}
$$

Solving for Δ,

$$
\begin{aligned}
\Delta &= \sigma_x^2 \left(\gamma^F\right)^2 \left(\sigma_2^2 + \sigma_1^2\right) + \left(\gamma^F \alpha + \left(1 - \gamma^F\right) \alpha^F\right)^2 \sigma_x^4 - \left(\gamma^F \alpha + \left(1 - \gamma^F\right) \alpha^F\right)^2 \sigma_x^4 \\
&= \sigma_x^2 \left(\gamma^F\right)^2 \left(\sigma_2^2 + \sigma_1^2\right). \qquad \text{(A.23)}
\end{aligned}
$$

Plugging Δ back to equation (A.22), we have

$$
\beta = \frac{\gamma^F \sigma_2^2 \sigma_x^2}{\sigma_x^2 \left(\gamma^F\right)^2 (\sigma_2^2 + \sigma_1^2)} = \frac{\sigma_2^2}{\gamma^F (\sigma_2^2 + \sigma_1^2)} = \frac{\gamma}{\gamma^F}
$$

Now solving for B,

$$
\begin{aligned}
\Delta B &= -\left(\gamma^F \alpha + \left(1 - \gamma^F\right) \alpha^F\right) \sigma_x^2 \left[\left(\gamma^F \alpha + \left(1 - \gamma^F\right) \alpha^F\right) \alpha \sigma_x^2 + \gamma^F \sigma_2^2\right] \\
&\quad + \alpha \sigma_x^2 \left(\left(\gamma^F\right)^2 \left(\sigma_2^2 + \sigma_1^2\right) + \left(\gamma^F \alpha + \left(1 - \gamma^F\right) \alpha^F\right)^2 \sigma_x^2\right) \\
&= -\left(\gamma^F \alpha + \left(1 - \gamma^F\right) \alpha^F\right) \sigma_x^2 \gamma^F \sigma_2^2 + \alpha \sigma_x^2 \left(\gamma^F\right)^2 \left(\sigma_2^2 + \sigma_1^2\right) \qquad \text{(A.24)}
\end{aligned}
$$

Plugging Δ back to equation (A.24), we have

$$
\begin{aligned}
B &= \frac{-\left(\gamma^F \alpha + \left(1 - \gamma^F\right) \alpha^F\right) \sigma_x^2 \gamma^F \sigma_2^2 + \alpha \sigma_x^2 \left(\gamma^F\right)^2 (\sigma_2^2 + \sigma_1^2)}{\sigma_x^2 \left(\gamma^F\right)^2 (\sigma_2^2 + \sigma_1^2)} = \alpha - \left(\alpha - \alpha^F + \frac{\alpha^F}{\gamma^F}\right) \gamma \\
&= (1 - \gamma) \alpha - \left(\frac{\gamma}{\gamma^F} - \gamma\right) \alpha^F
\end{aligned}
$$

as stated.

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}.

$$\underbrace{\pi_{t+3} - \mathbb{F}_t^{(\mu)}\left[\pi_{t+3}\right]}_{\text{Forecast Error}} = \alpha + \beta\left(\underbrace{\mathbb{F}_t^{(\mu)}\left[\pi_{t+3}\right] - \mathbb{F}_{t-1}^{(\mu)}\left[\pi_{t+3}\right]}_{\text{Forecast Revision}}\right) + \epsilon_{t+3} \tag{A.25}$$

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

$$\pi_{t+3} = 100 \times \left(\frac{P_t}{P_{t-1}} \times \frac{P_{t+1}}{P_t} \times \frac{P_{t+2}}{P_{t+1}} \times \frac{P_{t+3}}{P_{t+2}} - 1\right). \tag{A.26}$$

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.25) is constructed in the following way:

$$\begin{aligned}\pi_{t+3,t} - \mathbb{F}_t^{(\mu)}\left[\pi_{t+3,t}\right] \equiv 100 \times \Bigg[& \left(\frac{\pi_{t,t-1} - \mathbb{F}_t^{(\mu)}\left[\pi_{t,t-1}\right]}{400} + 1\right) \\ & \times \left(\frac{\pi_{t+1,t} - \mathbb{F}_t^{(\mu)}\left[\pi_{t+1,t}\right]}{400} + 1\right) \\ & \times \left(\frac{\pi_{t+2,t+1} - \mathbb{F}_t^{(\mu)}\left[\pi_{t+2,t+1}\right]}{400} + 1\right) \\ & \times \left(\frac{\pi_{t+3,t+2} - \mathbb{F}_t^{(\mu)}\left[\pi_{t+3,t+2}\right]}{400} + 1\right) - 1\Bigg]\end{aligned} \tag{A.27}$$

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

Table A.9: CG In-Sample Regressions of Forecast Errors on Forecast Revisions (Survey)

Regression: $\pi_{t+3,t} - \mathbb{F}_t[\pi_{t+3,t}] = \alpha + \beta(\mathbb{F}_t[\pi_{t+3,t}] - \mathbb{F}_{t-1}[\pi_{t+3,t}]) + \delta\pi_{t-1,t-2} + \epsilon_t$				
	(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)
$\pi_{t-1,t-2}$		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%.

In-sample analysis

Table A.9 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.10 presents the results from CG regressions when we replace the survey forecast with our machine forecast for SPF mean inflation. More specifically, we estimate is the following regression:

$$\underbrace{\pi_{t+3,t} - \mathbb{E}^{(\mu)}_{t+3|t}}_{\text{Machine Forecast Errors}} = \alpha + \beta\left(\underbrace{\mathbb{E}^{(\mu)}_t[\pi_{t+3,t}] - \mathbb{E}^{(\mu)}_{t-1}[\pi_{t+3,t}]}_{\text{Machine Forecast Revision}}\right) + \delta\pi_{t-1} + \epsilon_{jt+3}$$

where $\mathbb{E}^{(\mu)}_t[\pi_{t+3,t}]$ is the machine mean forecast made at time t and $\mathbb{E}^{(\mu)}_{t-1}[\pi_{t+3,t}]$ is the machine forecast made at time $t-1$.

Out-of-Sample Analysis

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

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

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}

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

Table A.10: CG Regressions of Forecast Errors on Forecast Revisions (Machine)

Regression: $\pi_{t+3,t} - \mathbb{E}_t[\pi_{t+3,t}] = \alpha + \beta(\mathbb{E}_t[\pi_{t+3,t}] - \mathbb{E}_{t-1}[\pi_{t+3,t}]) + \delta\pi_{t-1,t-2} + \epsilon_t$		
	(1)	(2)
Constant	-0.12	-0.13
t-stat	(-1.21)	(-0.94)
$\mathbb{E}_t[\pi_{t+3,t}] - \mathbb{E}_{t-1}[\pi_{t+3,t}]$	-0.04	-0.04
t-stat	(-0.22)	(-0.24)
$\pi_{t-1,t-2}$		0.00
t-stat		(0.08)
$\bar{R}^2$	0.0008	0.0008

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 $\mathbb{E}_t[\pi_{t+3,t}]$ is the machine mean forecast of annual inflation with information in period t and $\mathbb{E}_{t-1}[\pi_{t+3,t}]$ is the machine mean forecast of the same annual inflation but with information in $t-1$. 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%. The sample is 1995:Q1 to 2018:Q2.

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

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

Dynamic Responses to Cyclical Shocks–Local Projection

We follow Angeletos, Huo, and Sastry (2020) (AHS) and estimate the dynamic responses to inflation or GDP growth shocks from Angeletos, Collard, and Dellas (2018a) via local projection using a series of single equation regressions, one for each horizon $0 \leq h \leq H$ taking the form

$$z_{t+h} = \alpha_h + \beta_h\varepsilon_t + \gamma' W_t + u_{t+h} \tag{A.28}$$

where z_t is either the outcome variable at t, the survey forecast made at t, $\mathbb{F}_t^{(i)}[y_{j,t+h}]$, or the machine forecast made at time t, $\mathbb{E}_t^{(i)}[y_{j,t+h}]$. The dynamic responses plotted in the figures of the main text and below are given by the sequence of coefficients $\{\beta\}_{h=0}^{H}$, where W_t is a vector of control variables that are the same as those used in AHS and include one lag each of

Table A.11: Mean Square Errors for the CG Model and SPF

Forecast model: $\widehat{\pi}^{(\mu)}_{t+3} = \widehat{\alpha}^{(\mu)}_t + \left(1 + \widehat{\beta}^{(\mu)}_t\right) \mathbb{F}^{(\mu)}_t [\pi_{t+3}] - \widehat{\beta}^{(\mu)}_t \mathbb{F}^{(\mu)}_{t-1} [\pi_{t+3}]$			
	$MSE_{CG}/MSE_{\mathbb{F}}$		
Method	Quarterly Compound	Continuous Compound	CG Sample
Rolling 5 years	1.38	1.38	1.39
Rolling 10 years	1.29	1.29	1.29
Rolling 20 years	1.31	1.30	1.34
Recursive 5 years	1.69	1.68	1.71
Recursive 10 years	1.60	1.59	1.59
Recursive 20 years	1.33	1.30	1.34

Notes: The table reports the ratio of MSEs of the CG model forecast over the survey 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 sample spans the period 1969:Q1 - 2018:Q2. The CG sample refers to the sample in Coibion and Gorodnichenko (2015) that ends in 2014:Q4.

the outcome and survey forecast. We consider two outcome variables: inflation and real GDP growth. Following Angeletos, Huo, and Sastry (2020), we plot forecasts and outcome variables so that $\mathbb{F}^{(50)}_t [y_{j,t+h}]$ is lined up with $y_{j,t+h}$ along a vertical slice and the difference between the two is the forecast error. On the left-hand-side the forecasts are made at time t for period $t+h$, while the shock occurs at t. We compute the heteroskedasticity and autocorrelation robust (HAC) standard errors with a 4-quarter Bartlett kernel to calculate standard errors for the impulse responses. The ± 1 standard error bands are reported.

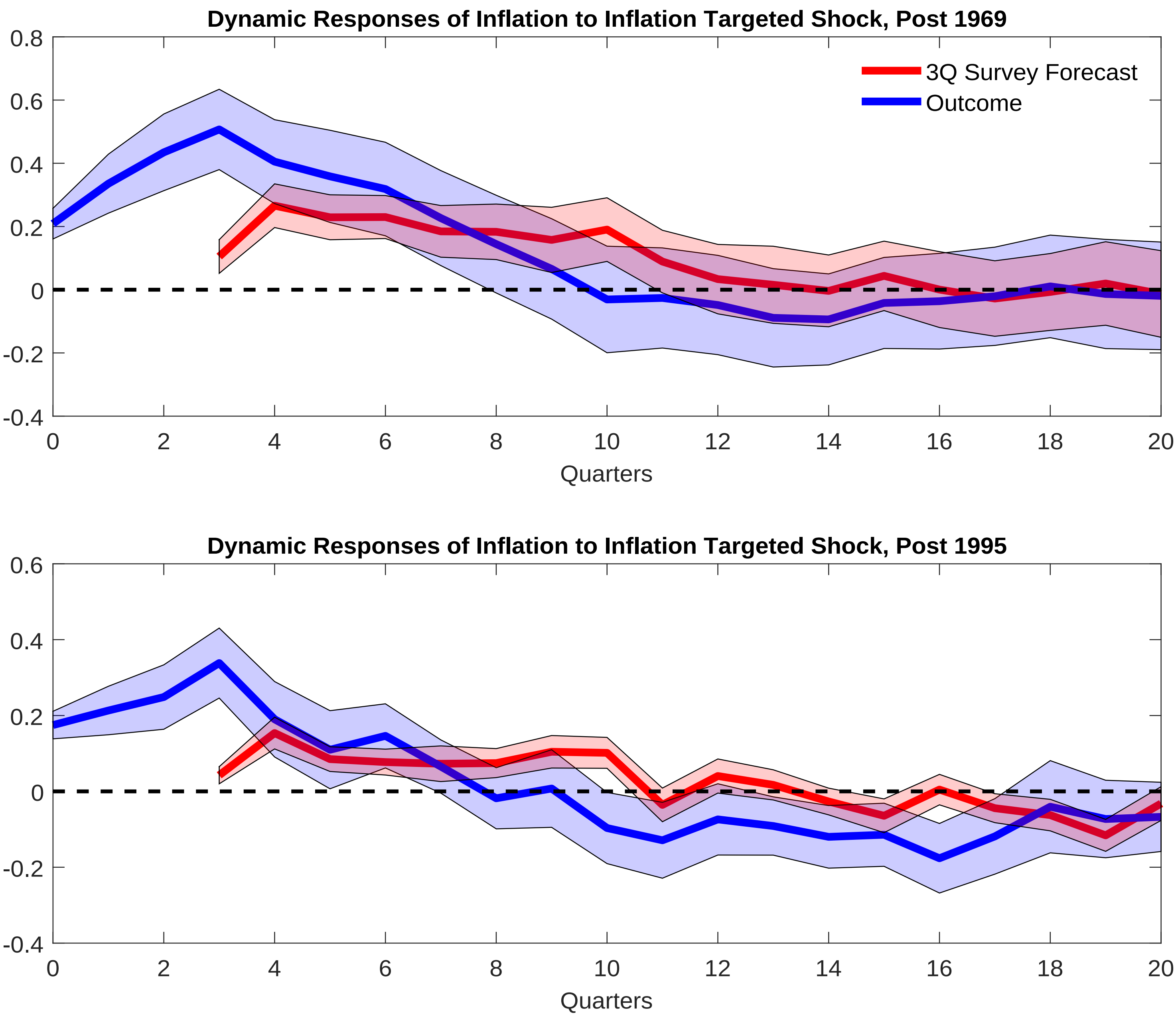
Top panel of Figure A.1 shows that we replicate the dynamic responses of inflation to an inflation targeted shock over the same sample used in Angeletos, Huo, and Sastry (2020). The bottom panel of Figure A.1 shows the dynamic responses are similar using the local projection estimation over our evaluation sample 1995:1-2018:Q2.

Machine Forecasts with Bloomberg Survey Data

To form an estimate of the median SPF machine forecast $\mathbb{E}^{(50)}_t [y_{j,t+h}]$ for four-quarter ahead GDP growth that does not use the median type's time t observation $\mathbb{F}^{(i)}_t [y_{j,t+h}]$, we can use similar professional forecasts from more timely survey data available prior to the time t SPF survey deadline. The Bloomberg (BBG) US consensus forecasts are updated daily (except for weekends and holidays). We use the median forecast from the Bloomberg Terminal for GDP on the closest day before the SPF survey deadline. The terminal reports daily quarter-over-quarter real GDP growth forecasts starting in 2003:Q1. To be consistent with the SPF forecasts, we construct the annual GDP growth forecast as follows.

Let $gY^{(Q/Q)}_{t+h}$ denote annualized quarter-over-quarter GDP growth in percent, h quarters ahead, and let $\mathbb{B}^{(50)}_t \left[gY^{(Q/Q)}_{t+h}\right]$ be the median BBG forecaster's prediction of this variable made

Figure A.1: Dynamic Responses: Forecast and Outcome



Dynamic responses of GDP and inflations. The shaded areas are 68% confidence intervals based on HAC standard errors with a Bartlett kernel and 4 lags. The x-axis denotes quarters from the shock. The outcome variable is inflation π_t and the shock is the inflation-targeted shock. The survey forecast is $\mathbb{F}_t^{(50)}[y_{t+3}]$. The shock time series are from Angeletos, Collard, and Dellas (2018a). In the first row, the impulse responses are estimated over sample 1969:Q1 to 2018:Q2. In the second row, the impulse responses are estimated over sample 1995:Q1 to 2018:Q2. In both rows, we "align" the forecast responses such that, at a given vertical slice of the plot, the outcome and forecast responses are measured over the same horizon, and the difference between the two is the forecast error. The vintage of observations on the outcome variable is final-release data.

at time t, where time t is the closest day before the SPF survey deadline listed in Table A.1. Bloomberg $\mathbb{B}_t^{(50)}\left[gY_{t+h}^{(Q/Q)}\right]$ are reported at annual rates in percentage points, so we convert to quarterly raw units before compounding. Let $y_{t+4,t}$ denote four-quarter real GDP growth. We construct the four-quarter real GDP growth BBG forecast from $gY_{t+h}^{(Q/Q)}$ as:

$$\mathbb{B}_t^{(50)}\left[y_{t+4,t}\right] = 100 \times ln\left(\prod_{h=1}^{4}\left(1+\frac{\mathbb{B}_t^{(50)}\left[gY_{t+h}^{(Q/Q)}\right]}{100}\right)^{\frac{1}{4}}\right).$$

$\mathbb{B}_t^{(50)}\left[y_{t+4,t}\right]$ exhibits a correlation with the SPF median annual GDP forecast $\mathbb{F}_t^{(50)}\left[y_{t+4,t}\right]$ of 96.4% over the common sample from 2003:Q1 to 2018:Q2.

To form an estimate of the median SPF machine forecast $\mathbb{E}_t^{(50)}\left[y_{j,t+h}\right]$ for four-quarter ahead GDP growth that does not use the median type's time t survey forecast $\mathbb{F}_t^{(i)}\left[y_{j,t+h}\right]$, we instead use the BBG professional consensus survey forecast which is publicly available on the closest day before time t SPF survey deadline. The estimation is the same as in the baseline estimation with two exceptions. First, we replace the survey deadline observation of the time t SPF median forecast series $\mathbb{F}_t^{(50)}\left[y_{j,t+h}\right]$ with the time t observation on the median forecast from the BBG survey, $\mathbb{B}_t^{(50)}\left[y_{t+4,t}\right]$. Second, the machine forecast is estimated over a shorter sample starting from 2003:Q1, when BBG data are available. The evaluation sample is for this estimation spans 2010:Q1-2018:Q2.

Table A.12: Results with Bloomberg Forecasts

ML: $y_{j,t+h} = \alpha_j^{(i)} + \beta_{jFt}^{(i)}\mathbb{F}_t^{(i)}\left[y_{j,t+h}\right] + B_{jZ}^{(i)} Z_{jt}\epsilon_{jt+h}$		
SPF GDP Median Forecast		
Replace survey deadline observation $\mathbb{F}_t^{(50)}$ with $\mathbb{B}_t^{(50)}$		
	Use $\mathbb{F}_t^{(50)}$ for all t	Replace survey deadline $\mathbb{F}_t^{(50)}$ with $\mathbb{B}_t^{(50)}$
MSE_E/MSE_F	0.81	0.85
OOS R^2	0.19	0.15

Notes: This table reports the MSE ratios with and without using Bloomberg consensus forecasts. The second column reports the results when SPF median forecasts are used for all quarters. The second column reports the results when the current-quarter SPF median forecast of GDP growth is replaced by the Bloomberg median forecast and include one lag of SPF forecasts of all types. $MSE_{\mathbb{E}}$ and $MSE_{\mathbb{F}}$ denote the machine and SPF survey mean-squared-forecast-errors, respectively, for 4-quarter-ahead forecasts, averaged over the evaluation sample. The out-of-sample Rsquared, OOS R^2, is defined as 1-$MSE_{\mathbb{E}}/MSE_{\mathbb{F}}$. The evaluation period is 2010:Q1 to 2018:Q2;