NBER WORKING PAPER SERIES

THE PRESTAKES OF STOCK MARKET INVESTING

Francesco Bianchi
Do Q. Lee
Sydney C. Ludvigson
Sai Ma

Working Paper 34420
http://www.nber.org/papers/w34420

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 2025

Ludvigson is grateful to the CV Starr Center for Applied Economics at NYU for financial support. The views expressed are those of the authors and do not necessarily reflect those of the Federal Reserve Board, the Federal Reserve System, or the National Bureau of Economic Research.

The Prestakes of Stock Market Investing
Francesco Bianchi, Do Q. Lee, Sydney C. Ludvigson, and Sai Ma
NBER Working Paper No. 34420
October 2025
JEL No. G1, G17, G40, G41

ABSTRACT

How rational is the stock market and how efficiently does it process information? We use machine learning to establish a practical measure of rational and efficient expectation formation while identifying distortions and inefficiencies in the subjective beliefs of market participants. The algorithm independently learns, stays attentive to fundamentals, credit risk, and sentiment, and makes abrupt course-corrections at critical junctures. By contrast, the subjective beliefs of investors, professionals, and equity analysts do little of this and instead contain predictable mistakes–prestakes–that are especially prevalent in times of market turbulence. Trading schemes that bet against prestakes deliver defensive strategies with large CAPM and Fama-French 5-factor alphas.

Francesco Bianchi
Johns Hopkins University
Department of Economics
and CEPR
and also NBER
francesco.bianchi@jhu.edu

Do Q. Lee
New York University
Department of Economics
dql204@nyu.edu

Sydney C. Ludvigson
New York University
Department of Economics
and NBER
sydney.ludvigson@nyu.edu

Sai Ma
Federal Reserve Board of Governors
sai.ma@frb.gov

1 Introduction

"Prestakes" refer to predictable mistakes in financial markets. The absence of predictable mistakes in investor beliefs is arguably the most important defining feature of a rational and efficient stock market. While there are good reasons to expect that the stock market is neither perfectly rational nor perfectly efficient at processing information, quantifying the extent to which this is so has been a fundamental challenge of asset pricing research. So far, empirical work has largely proceeded by investigating whether forecast errors made by survey respondents deviate from the canonical standard of full information and rational expectations (FIRE). Yet knowing whether a deviation exists is not the same as quantifying its importance for the subjective beliefs of investors. For this we need an explicit measure of non-distorted and efficient expectation formation, i.e., a measure of objective beliefs, with which to compare the subjective beliefs of investors.

In this paper we form one such measure for equity market participants, and in so doing provide new estimates of the extent to which the stock market deviates from a practical standard of non-distorted and efficient expectation formation. For brevity, we refer to any such deviation–whether due to non-rational expectations or informational inefficiency–as a *belief distortion*.

Our approach begins with the observation that modern-day market participants have access to vast information sets and advanced computing resources and algorithms, but rests on the premise that their beliefs could still be distorted by behavioral biases, inefficiencies in information processing, slow learning, and/or sub-optimal attentiveness to information. A belief distortion in our framework is a *predictable* mistake, one that arises from a demonstrable misuse of available information. Predictable mistakes are distinct from ex post forecast errors. They result instead from a sustained pattern of ex ante misjudgment. For investors, such prestakes are especially concerning when they happen in times of heightened market volatility, since that is when the stakes are highest.

In developing a practical measure of objective beliefs, we argue that three elements are essential. First, any such measure must mind key features of real-world expectation formation in financial markets, including the data-rich setting in which decisions are made and the out-of-sample nature of predictions. Failure to do so by (for example) comparing the subjective beliefs of investors with forecasts from low-dimensional prediction models that we understand to work well only with hindsight can lead to erroneous conclusions about belief distortions and their relation to financial markets. Second, the measure should remove the human element from decision-making as much as possible. This is important in order to strip out the influence of behavioral bias and/or cognitive limits to efficient information processing. Third, the measure should be capable of efficiently processing large quantities of information, an obvious requirement for any practical notion of informational efficiency.

To accommodate all of these ingredients, we develop a machine learning algorithm to independently make a sequence of live stock market predictions in a data-rich setting. The machine is trained to adapt an artificial neural network to evidence of a changing economic landscape, learning about stock market market volatility over an initial sample that spans 1970 to 2005, a period that includes several major boom/bust episodes. We structure the algorithm so that the machine's forecasts can differ from the survey forecasts only if the machine finds evidence of predictable mistakes in the survey responses immediately prior to the machine making a true out-of-sample forecast. The algorithm is then tested repeatedly over the remaining sample that ends in 2023. Importantly, the machine is required to do its analysis without look-ahead advantages, using inputs that consist only of historical, time-stamped data that we compile ourselves and can verify would have been available to market participants when making a prediction.[1]

The resulting algorithm produces a sequence of machine "beliefs" against which the measured expectations of market participants may be assessed. For measured expectations about stock market returns, we use survey data on individuals, chief financial officers (CFOs), and professional forecasters that have been widely used in the literature, and convert them into measures of subjective risk premia. For expectations about corporate earnings, we use surveys of professional equity market analysts from IBES, forming a bottom-up aggregation to

[1]In particular, we do not use commercial large language models trained on contemporary data.

the level of the S&P 500.

We focus on three investigations central to questions on the rationality and efficiency of the U.S. stock market. First, we investigate whether survey respondents make predictable mistakes in their forecasts of stock market returns and corporate earnings for the U.S. aggregate stock market. We find that they do, and investigate why. The magnitude of these prestakes is evident from the machine's superior forecasting performance over an extended testing sample from 2005:Q1-2023:Q4, as compared to investors, CFOs, professional forecasters, and earnings analysts. For predicting one-year ahead stock market returns, the machine algorithm is at least 30% more accurate than investors, CFOs, and professional forecasters who–unlike the machine–failed to forecast negative returns in the 2008 global financial crisis (GFC) and in the bear market of 2022 that coincided with the spike in general price inflation. For predicting one-year ahead corporate earnings growth, the machine is 64% more accurate than professional equity analysts over the full testing sample and 70% more accurate over a period that includes the GFC and its aftermath.

Since the machine forecasts are based on information that we verify would have been available to survey respondents but are systematically more accurate, the survey forecasts must implicitly mis-weight or entirely ignore relevant information employed by the machine that could have been used to improve survey forecast accuracy, but evidently was not. It follows that our results showing a substantial forecast accuracy gap are isomorphic to directly finding that the time t survey forecast errors are correlated with information available to survey respondents at the time of their forecasts, a common approach to identifying departures from FIRE without quantifying their importance. The distinction with our approach arises because we have a measure of objective beliefs, which permits us to quantify the magnitude of any systematic errors in survey expectations by a simple comparison with objective expectations. We refer to any non-zero gap between the two at time t as the time-t *bias* and use it as a period-specific measure of the magnitude of prestakes, i.e., the systematic distortion in subjective beliefs.

To study the reasons for these predictable mistakes, we ask the machine to learn about the forecasting models that best explain the survey responses at each point in time. An analysis of this output reveals three general ways in which subjective beliefs differ from objective beliefs. First, subjective beliefs vary too little and do not adapt quickly enough and non-linearly enough to evidence of a changing economic landscape. Second, we find little evidence that survey respondents learn from past mistakes. Third, survey responses exhibit a form recency bias with forecasts that are largely explained by a local mean in the predicted variable, even in times of extreme market turbulence. In contrast to each of these, the machine often accesses its long-memory functionality by paying greater attention to credit risk variables, fed funds futures market surprises, consumer spending, and media-sentiment variables (among others) to course-correct and identify turning points, especially in turbulent times. It is at these times that the machine identifies the largest distortions in measured beliefs.

Meaningful distortions in subjective beliefs raise the possibility that real-world investors have a skewed view of the objective tradeoff between risk and reward. This brings us to our second investigation, which compares the risk-return tradeoff in subjective beliefs to measures of the same in objective beliefs. For subjective beliefs, we use our measures of subjective risk premia from return surveys along with several measures of subjective risk perceptions from the extant literature. For objective beliefs, we use our machine-based measures of risk premia along with a measure of objective financial market uncertainty from the extant literature. Although we find a positive risk-return tradeoff in both subjective and objective beliefs, objective risk premia move between 2 and 6.8 times more with objective measures of risk than subjective risk premia move with subjective measures of risk. This suggests that subjective risk premia are sub-optimally responsive to evidence of changing risk, even if we allow for the measures that investors actually perceive.

For our third investigation, we ask whether strategies that bet on our measure of objective beliefs earn positive risk-adjusted returns, or "alpha." Since the machine forecasts can deviate from the survey forecasts only if the algorithm detects predictable mistakes in subjective beliefs, strategies that explicitly bet on objective beliefs are implicitly betting against the prestakes in subjective beliefs. We find that trading on our objective expectations sequence could have been used to construct defensive strategies with large CAPM and Fama-French 5-factor alphas. A strategy that would have bought the CRSP index when the machine excess return forecast was positive and otherwise invested in a Treasury bill earns CAPM and Fama-French 5-factor alphas of 4.6 and 4.4 percent per annum, respectively. A strategy that would have bought the CRSP index when the machine excess return forecast was positive, and otherwise invested in the Treasury-bill while shorting the market when the machine's raw return forecast was negative, earns CAPM and Fama-French 5-factor alphas of 9 and 8 percent annum, respectively. These large machine alphas are "earned" almost entirely in and around periods of market turbulence, such as the GFC and the bear market of 2022, when the strategies were relatively successful at mitigating losses. By contrast, strategies based on the subjective beliefs of market participants all earn zero or negative alphas because they virtually never anticipate losses. While not a definitive test, we view these results as suggestive of a connection between prestakes in subjective beliefs and realized stock market dynamics. After all, if our measured subjective beliefs and their identified distortions are insufficiently representative of "market" beliefs and "market" distortions, we would not expect trading on the basis of the identified distortions to be profitable.

It is tempting to ask whether the forecasting power of our objective belief sequence is large or small in some absolute sense. Such questions ignore the key context of our analysis, namely the unpredictable shocks that drive most of stock market variation. Without perfect foresight, there is no algorithm capable of error-free ex ante forecasts, a universal law that is especially relevant for low signal-to-noise variables such as financial returns. The machine algorithm we study is not immune to this law, and it–like our survey respondents–sometimes makes large ex post mistakes. This is evident by comparison with the infeasible perfect foresight versions of the trading strategies just discussed, where alphas for the second strategy take on much larger values than versions based on objective beliefs, in the range of 14%. For this reason, statements on forecasting power in our context are always relative statements, where "success" can only be measured by comparison with other, truly ex ante, predictions.

We argue that ex post mistakes–as distinct from predictable mistakes–underscore the role of largely unforeseen events and/or hard-to-learn structural change in generating occasionally large prediction error, not all of which can be attributed to a systematic bias in expectations.

Relation to the Literature In developing our approach to quantifying belief distortions in the stock market, we draw on a great deal of related literature. Methodologically, our approach to using machine learning as a means of measuring the magnitude of ex ante expectational errors in survey forecasts builds closely on the approach in Bianchi, Ludvigson and Ma (2022) (BLM hereafter). BLM focuses on biases in survey expectations of GDP growth and inflation, while this paper is focused on the stock market. Yet the present paper differs from BLM in a number of ways that go well beyond this distinction. This includes our use of different, higher dimensional specifications more suitable for financial markets, the training of the machine to independently adapt its network to evidence of economic change, and our use of machine learning to predict the forecasting models of survey respondents, which allows for a direct comparison of how and why objective and subjective forecasts differ over time.

A large and growing body of literature studies data on subjective beliefs in financial economics. A number of these find evidence consistent with overreaction in stock market expectations (Barberis, Shleifer and Vishny (1998), Chen, Da and Zhao (2013), Bordalo, Gennaioli and Shleifer (2018), Bordalo, Gennaioli, Ma and Shleifer (2020), Bordalo, Gennaioli, LaPorta and Shleifer (2019), Nagel and Xu (2022), Afrouzi, Kwon, Landier, Ma and Thesmar (2023), Bordalo, Gennaioli, LaPorta and Shleifer (2022), De La O and Meyers (2021, 2023) Hillenbrand and McCarthy (2021).) Other researchers have argued that investors may be rationally or behaviorally inattentive due to information processing constraints or cognitive biases, thus underreacting to information (e.g., Mankiw and Reis (2002), Woodford (2002), Sims (2003), Gabaix (2019), Kohlhas and Walther (2021)). In this literature, empirical investigations usually ask whether measured subjective beliefs exhibit evidence of departures from FIRE. A common approach is to run in-sample regressions of survey forecast errors on lagged forecast revisions.

We build on and extend this literature by offering two key innovations. First, we go beyond using data on subjective beliefs and documenting violations of FIRE to quantifying their importance in a real-world, real-time context. Since there could be many different such violations–including some that have never been discovered–quantifying their cumulative importance requires an explicit measure of objective beliefs with which to compare to subjective beliefs. Moreover, placing distortions in a real-world context requires a practical measure of objective beliefs rather than a model-based benchmark. Taken literally, FIRE and the models in which they are embedded are theoretically convenient but wholly unrealistic standards that would require the models to be specified without error, while market participants must know the true data generating processes and quickly learn its parameters in the wake of any kind of structural change. Our approach instead seeks a practical and less theoretical baseline measure of objective beliefs relevant for investors today. Second, we discipline our analysis in a number of specific ways: (i) we rely exclusively on information that we can verify could have been known to survey respondents on or before the survey response deadline, (ii) we require the machine to independently choose every aspect of the forecasting specification ex ante rather than ex post, (iii) we require our benchmark to learn about and test for structural change adaptively and with experimentation, and (iv) we evaluate distortions using true out-of-sample prediction in the external validation step. The strict adherence to these principles is important to avoid overstating departures from FIRE in real-world decision making.

Our use of machine methods to learn about the forecasting models that best explain survey responses connects with recent work by Bastianello, H Décaire and Guenzel (2024), who build a mental model of sell-side analysts' forecasts that is informed by their written reports. While our empirical tools differ, we view them as complimentary approaches in the quest toward a comprehensive understanding of subjective beliefs. Despite the differing methodologies, a similarity in our findings is that equity analyst forecasts rely on few inputs and stick to those inputs over time with little adaptation.

Our study also draws on machine learning insights from previous work, especially the tools used in predicting financial returns, e.g., Gu, Kelly and Xiu (2020), Bybee, Kelly, Manela and Xiu (2021), Cong, Tang, Wang and Zhang (2021), and Nagel (2021). These studies are focused primarily on showing the power of supervised learning algorithms for asset return prediction. While our algorithms utilize supervised learning, they differ from these studies in that they are specifically designed to uncover and quantify distortions in subjective beliefs.

The rest of this paper is organized as follows. Section 2 describes our empirical framework and data. Section 3 presents our main results, and Section 4 concludes. A large amount of additional information is reported in an Online Appendix.

2 Framework

This section highlights the key aspects of our approach to using machine learning for measuring the systematic expectational errors embedded in human judgments, i.e., their predictable mistakes. To conserve space, we provide an abbreviated description of many technical aspects of our approach here, referring the interested to BLM and the Online Appendix for further detail.

2.1 Machine Learning and Data

Machine Model We are interested in forming a time t objective machine expectation of a generic variable $y_{j,t+h}$ indexed by j whose $h \geq 1$ period ahead value the machine is asked to predict. In our case the target variable $y_{j,t+h}$ will be either corporate earnings growth or stock market returns. The machine specification is estimated over rolling sub-samples and is written generically as:

$$y_{j,t+h} = G\left(\mathcal{X}^t, \boldsymbol{\beta}_{j,h,t}\right) + \epsilon_{jt+h}. \tag{1}$$

where $\mathcal{X}^t$ denotes the history of data inputs from time $1, ..., t$ available in real time (including an intercept), and $G(\cdot)$ is a machine learning estimator that can be represented by a high dimensional set of finite-valued parameters $\boldsymbol{\beta}_{j,h,t}$. Our main approach uses the Long Short-Term Memory (LSTM) deep sequence recurrent neural network for $G\left(\mathcal{X}^t, \boldsymbol{\beta}_{j,h,t}\right)$.[2] This particular specification is useful because it allows the machine to learn about whether to draw on memory that is more long-term or near-term in nature when making predictions over time. Note that the time t subscript on $\boldsymbol{\beta}_{j,h,t}$ is used to denote one in a sequence of time-invariant parameter estimates obtained from rolling sub-samples, rather than estimates that vary over time within a sample. Estimation of (1) delivers a time t machine "belief" about $y_{j,t+h}$ given by the estimated value of $G\left(\mathcal{X}^t, \boldsymbol{\beta}_{j,h,t}\right)$. We denote this time t objective belief as $\mathbb{E}_t\left[y_{j,t+h}\right]$.

With this estimator in hand, we build off of the five step algorithmic approach of BLM: 1) Sample partitioning. At time t, a prior sample of size $\dot{T}$ is partitioned into two sub-sample windows: a "training" sub-sample consisting of the first T_E observations, and a hold-out "validation" sample of T_V subsequent observations, with $\dot{T} = T_E + T_V$. 2) Training. The training sample is used to estimate parameters subject to a specific set of hyperparameters, a specific network architecture, and specific window lengths, T_E and T_V. This process is repeated for multiple architectures and hyperparameters. 3) Hyperparameter tuning and re-optimization. A "validation" sample is used for tuning hyperparameters, selection of network architecture, optimization of training and validation window lengths, and checking for predictable mistakes in a paired-survey forecast (see below). In this step, pseudo out-of-sample forecasts of $y_{j,t+h}$ in the validation sample are computed based on the estimated models in the training sample. The algorithm calculates the value of the objective function based on the forecast errors in the validation sample, and iteratively searches for hyperparameters, window lengths, and network architectures that optimize the validation objective. At each step, the model is re-estimated from the training data subject to the prevailing set of hyperparameters, win-

[2] The LSTM estimator with N hidden layers $h_t^N \in \mathbb{R}^{D_{h^N}}$ takes the form

$$G^{LSTM}\left(\mathcal{X}^t, \beta_{jh}\right) = \underbrace{W^{(yh^N)}}_{1\times D_{h^N}} \underbrace{h_t^N\left(\mathcal{X}^t, \beta_{jh}\right)}_{D_{h^N}\times 1} + \underbrace{b_y}_{1\times 1}.$$

See the Online Appendix for the detailed specification.

dow lengths, and network architecture structure. 4) Out-of-sample prediction. In this step a truly out-of-sample prediction is made, with each iteration's prediction saved as part of an extended "testing" sample. 5) Roll forward and repeat. The end product is a sequence of time t objective beliefs about $y_{j,t+h}$, denoted $\{\mathbb{E}_t [y_{j,t+h}], \mathbb{E}_{t+1} [y_{j,t+1+h}], ..., \mathbb{E}_t [y_{j,T}]\}$.

Several points about the algorithm bear emphasizing. First, the algorithm is run multiple times while being "paired" with a different survey forecast in each case. We structure the algorithm so that the machine forecast selected from step 3 can only differ from the paired-survey forecast if it demonstrably improves upon the survey forecast in pseudo out-of-sample predictive accuracy over the multi-year hold-out validation sample immediately prior to making a true out-of-sample forecast in step 4. Bearing in mind that these pseudo out-of-sample forecasts rely only on information we verify would have been available to the survey respondent, this step ensures that the machine forecast can only differ from the survey forecast if the machine finds predictable mistakes in the survey response. Second, following BLM, each iteration renews the optimized selection of training and validation sample windows lengths, allowing the machine to adaptively choose how much of the historical data to use when estimating parameters and determining the depth and width of neural network architecture structure. Steps (1)–(3) are repeated for new values of $T_E \in \{\underline{T}_E, \ldots, \bar{T}_E\}$ $T_V \in \{\underline{T}_V, \ldots, \bar{T}_V\}$ such that alternative partitions satisfy $T_E + T_V \leq \dot{T}$, where shorter window lengths remove consecutive observations at the start of the prior sample rather than at the end. This is especially useful for earnings growth, which, like output growth, is subject to slower moving structural shifts or outright breaks, implying that no single set of window lengths is likely to work best in all time periods. Stock market returns are more stationary, making it optimal to set training sample window lengths to span long fixed or expanding time periods in order to take advantage of the machine's long-memory features.

We compute mean-squared forecast errors for the machine relative to survey benchmarks by averaging across the sequence of squared forecast errors in the testing sample for $t = (\dot{T} + h), \ldots, T$. The accumulation of true out-of-sample forecast errors serves as an external validation step that exists outside of the optimization loop. It is crucial that the testing sample be large enough so that the evidence on relative forecast accuracy is not the result of chance. We therefore require that our testing sample cover an extended time-frame from 2005:Q1 to 2023:Q3. At the same time, we need to reserve a minimum number of past observations for training and hyper-parameter tuning. Putting these two considerations together, we choose a historical sample that begins in 1970:Q1, which allows the machine to contemplate combined forecast and validation samples as long as 140 quarters.

As mentioned, the machine forecast can only differ from the paired-survey forecast if it finds evidence of predictable mistakes in the survey response. It is important emphasize that this does not mean the machine will always outperform the survey. Indeed, the machine can uncover a predictable mistake and yet under-perform by chance, even if it outperforms on average. Since these represent instances of ex post bad luck rather than ex ante bad judgment, we do not count them as predictable mistakes. It follows that we can quantify predictable mistakes in survey expectations by a simple comparison with the machine forecast. We refer to the time-varying gap between the two as the time t bias, which is simply the difference between a subjective investor forecast from survey s, $\mathbb{F}_{s,t}[\cdot]$, and the machine forecast of the same object $\mathbb{E}_t[\cdot]$, i.e., $Bias_{s,t}[y_{t+h}] \equiv \mathbb{F}_{s,t}[y_{t+h}] - \mathbb{E}_t[y_{t+h}]$. In particular, bias in expectations is measured relative to the machine forecast, not relative to an ex post outcome. $Bias_{s,t}[\cdot]$ is our period-by-period measure of prestakes in subjective beliefs. Note that any instance of $Bias_{s,t}[\cdot] \neq 0$ represents a systematic error in subjective beliefs, with $Bias_{s,t}[\cdot] > 0$ indicating excessive optimism about earnings growth or returns, and $Bias_{s,t}[\cdot] < 0$ indicating excessive pessimism.

Data and Information Processing The meta data-set used for this project consists of thousands of real-time economic time series and spans the period 1970:Q1 through 2023:Q4. These inputs consist of data with daily to biannual sampling intervals. In what follows, we provide a brief summary of the data and how it is used. A complete description of the data, sources, and mapping to the model is provided in the Online Appendix.

Many series in the machine's input layer are converted to diffusion index factors before being passed to the machine estimator. 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] The diffusion index aspect of our methodology is standard, so we cover this step in the online Appendix.

The initial set of data inputs include a real-time macro dataset of 92 indicators, a panel dataset of 147 monthly financial indicators, and daily "up-to-the-forecast" financial market variables from five broad classes of financial assets: (i) commodities prices (ii) corporate risk variables including credit spreads (iii) equities (iv) foreign exchange, and (v) government securities. Hundreds of other inputs are used at mixed-sampling intervals, including consensus forecast surprises for macro data releases, stock market jumps around news events, jumps in Fed funds futures around FOMC releases, and daily text-based factors estimated by Latent Dirichlet Allocation (LDA) analysis from around one million articles published in the *Wall Street Journal* between January 1984 to June 2024.[4]

We find it useful to separate the large number of inputs as falling into one of three general categories: "macro fundamentals," "non-fundamentals," and "media-sentiment." Macro fundamentals include data or factors that capture variation in employment, hours, output, corporate earnings, the general price level, and short-term interest rates that include federal funds futures and Treasury notes dated 3 months or less. Short rates are included in this category as a risk-free component of the discount rate, since they are largely under control of the central bank. This category also includes a very large number of data "surprises"– deviations of data releases from consensus forecasts or jumps in fed funds futures around central bank announcements. Non-fundamental data include all longer-dated financial market series from bond, stock, and other financial markets. These variables are classified as

[3]It is straightforward 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. See BLM for further discussion.

[4]Following Bybee et al. (2021) the LDA factor estimation is based on a randomly selected sub-sample of 200,000 articles over the same period.

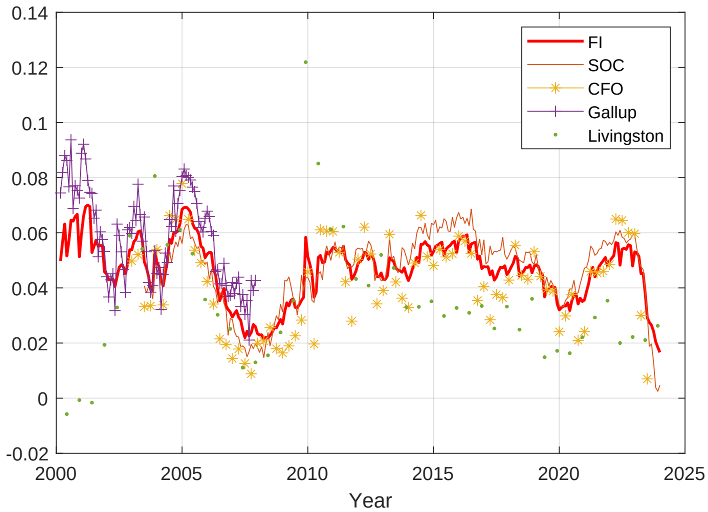
non-fundamentals because their longer duration mean they are influenced by both subjective beliefs and possible behavioral bias, as well as expected fundamentals. Media-sentiment data are constructed using natural language processing techniques. Following the procedure of Bybee et al. (2021), we estimate real-time LDA factors from Wall Street Journal article texts. We use them as a crude measure of media sentiment, since their role in the machine forecasts is always measured after controlling for macro fundamentals and non-fundamentals thereby isolating the additional information in news articles. For a complete list of variables included in these categories see Table A.7 in the Online Appendix.

To measure realized stock market returns, we use the annual Center for Research in Securities Prices value-weighted return series (CRSP-VW, includes dividends). For survey expectations of returns, we use mean forecasts from UBS/Gallup, the Michigan Survey of Consumers (SOC), the Conference Board (CB), the CFO Survey from the Richmond Federal Reserve Bank, and the Livingston survey. Surveys vary according to whether they ask about returns or price growth. For surveys that ask about price growth expectations, we adjust these with the dividend yield to arrive at a subjective return expectation. For each subjective one-year ahead expected stock market return series, we construct the subjective risk premium by subtracting the 1-year Treasury bill (T-bill) yield that prevailed at the time of the survey.[5] We also form a measure of investor expectations as a whole by treating each of these surveys as a noisy signal of the true underlying return expectations of investors using a state space estimation (see Online Appendix 3.4 for details). This also serves to create a single time series of investor expectations, which is useful since not all of the different return expectation surveys are available over the same time periods or at the same frequencies. The state-space estimation uses different measurement equations to include observations for series when they are available and exclude them when they are missing. We refer to this single series as the "filtered investor" (FI) series for brevity. The resulting FI series is plotted in Figure 1 along with individual survey measures. The FI series extracts the common signal from the various

[5]The SOC asks about the probability of a rise in the market and the CB asks for a categorial belief on whether the market will rise, stay the same, or fall. The Appendix explains how we convert these into a quantitative point forecast.

noisy surveys and provides in a long time series of subjective investor beliefs.

Figure 1: Filtered Investor Expected Return Series (FI)



Notes: Figure plots 1-year-ahead filtered investor return expectations (FI) extracted using a state-space model that combines multiple investor forecasts (Gallup/UBS, CFO, Livingston, and SOC) mapped to an AR(1) latent factor. The estimation uses the Kalman filter to fill in missing observations and idiosyncratic measurement error. Expectations are expressed in terms of total stock returns including dividends. Survey inputs for Livingston and SOC surveys are converted into forecasts of stock returns (with dividends) by adjusting for the dividend yield following Nagel and Xu (2022).

We measure corporate earnings for the S&P 500 with IBES "street earnings" and take this as the forecast target for IBES analysts. Street earnings differs from GAAP earnings by excluding discontinued operations, extraordinary charges, and other non-operating items.[6] To measure subjective expectations of earnings growth, we use data from IBES on the median forecast of professional equity analysts, converting their forecasts of earnings-per-share at the firm level to forecasts of earnings growth rates and finally build bottom-up measures of actual and subjectively expected aggregate S&P 500 earnings growth using the constituents of the S&P 500 at each point in time (see Appendix 3.4 for details).

[6]According to the IBES user guide, analysts submit forecasts after backing out these transitory components, and IBES constructs the realized series to align with those forecasts. While analysts have some discretion over which items to exclude, Hillenbrand and McCarthy (2024) demonstrate that the target of these forecasts corresponds closely to earnings before special items in Compustat, suggesting that street earnings accurately reflect the measure analysts are targeting.

3 Results

This section presents our estimation results. We begin with preliminary analysis of machine forecasting performance as compared to surveys and other benchmarks.

3.1 Preliminary Analysis

Table 1 summarizes the forecasting performance of the machine relative to a survey or other benchmark for stock market returns or corporate earnings growth for the S&P 500. The forecast horizon in most cases is the next four quarters, i.e., 1-year-ahead forecasts, with exceptions discussed below. Relative forecasting performance is reported as the ratio of the mean-square-forecast-error (MSE) of the machine, "$\text{MSE}_{\mathbb{E}}$" to that of a paired-survey forecast or other benchmark, and denoted "$\text{MSE}_{\mathbb{F}}$", where the sample for these comparisons largely depends on the availability of the survey data. In considering the results, it's important to bear in mind that the machine forecast can only differ from a paired-survey forecast if the algorithm concludes the survey is making predictable mistakes in the hold-out sample immediately prior to making a true out-of-sample forecast. This implies that the only way the machine can improve upon the forecast accuracy of a survey is if it concludes the survey response is making a predictable mistake.

Panel (a) of Table 1 considers forecasts of stock market returns over the next four quarters. Comparisons between machine and investor forecasts are made for five surveys of the stock market, plus our FI series: the Gallup/UBS poll of corporate executives, the CFO survey of corporate executives, and the Livingston Survey of forecast professionals, the Conference Board (CB), and the University of Michigan's Survey of Consumers (SOC). The subsamples for comparison depend on the availability of the survey data. The Gallup/UBS survey is only available for the short time period of 2005:Q1-2008:Q4, but the CFO and the FI series are available from 2005:Q1 to the end of our sample in 2023:Q4. We also compare the machine performance to a recursive estimate of the historical mean, given findings elsewhere that it is difficult for statistical models to beat the historical mean when predicting excess stock market returns out-of-sample (Goyal and Welch (2008)). For comparisons with the historical mean, we also consider forecast horizons from 8 to 20 quarters ahead. Panel (a) shows that the machine algorithm produces forecasts of excess returns that are substantially more accurate than all of these benchmarks. The ratio of the machine-MSE to survey-MSE is 0.68, 0.67, 0.74, 0.65, and 0.75 when compared to the CB, SOC, Gallup/UBS, CFO, and LIV surveys, respectively. This corresponds to out-of-sample R^2 statistics of 0.32, 0.33, 0.26, 0.35, and 0.25, respectively, implying that the machine forecasts are 32%, 33%, 26%, 35%, and 25% more accurate than the respective investor forecasts. The machine is also substantially more accurate than the historical mean, with the ratio of the machine MSE to historical mean MSE equal to 0.64, 0.69, 0.55 and 0.59 for one-year, two-year, and three-year excess return forecasts.

Table 1, panel (b), considers forecasts of earnings growth on the S&P 500. We compare the machine expectations to the median IBES equity analyst forecasts of earnings growth over the next four quarters, i.e., 1-year-ahead earnings growth on the S&P 500. We also consider the IBES long-term growth forecasts (LTG). Here we follow Bordalo et al. (2019) in aggregating the value-weighted firm-level long-term growth forecasts of the median analyst to obtain LTG at the S&P 500 level. Since the wording of the LTG question does not state the precise horizon over which analysts are asked to report expectations of "long-term" growth, we consider two possibilities for the purpose of evaluating forecast errors. First, we interpret LTG as measuring annual five-year forward growth expectations, i.e., annual earnings growth from four to five years ahead. Alternatively, we interpret LTG as measuring the annualized nine-year growth rate one year forward, i.e., annualized earnings growth from one to 10 years ahead.[7] The first interpretation is motivated by Bianchi, Ludvigson and Ma (2024), who compared the mean squared forecast errors using LTG to predict different

[7]For the LTG question, IBES states that "The long term growth rate represents an expected annual increase in operating earnings over the company's next full business cycle. These forecasts refer to a period of between three and five years." Interpreting LTG as an expected annual n-year forward growth rate (rather than the expected annualized n-year growth rate), is consistent with the reference to the *next* full business cycle and makes the stable median LTG forecast easier to reconcile with the volatile median one-year growth forecast.

Table 1: Forecasting Performance: Machine v.s. Benchmarks

Mean-square-errors of machine forecasts ($\text{MSE}_{\boldsymbol{E}}$) versus survey/benchmark ($\text{MSE}_{\mathbb{F}}$)						
(a) Stock Returns						
Horizon h (quarters)	4	4	4	4	4	4
Survey/benchmark	CB	SOC	Gallup/UBS	CFO	Livingston	FI
$\text{MSE}_{\mathbb{E}}/\text{MSE}_{\mathbb{F}}$ (LSTM)	0.684	0.666	0.738	0.646	0.745	0.657
OOS R^2	0.316	0.334	0.262	0.354	0.255	0.343
w^*	1.000	1.000	1.000	1.000	0.972	1.000
$\text{MSE}_{\mathbb{F}}/\text{MSE}_{\mathbb{H}}$ (Mean)	0.937	0.963	1.019	0.992	0.914	0.975
Horizon h (quarters)	4	8	12	16	20	
Survey/benchmark	Mean	Mean	Mean	Mean	Mean	
$\text{MSE}_{\mathbb{E}}/\text{MSE}_{\mathbb{F}}$ (LSTM)	0.641	0.686	0.547	0.588	0.719	
OOS R^2	0.359	0.314	0.453	0.412	0.281	
w^*	1.000	1.000	1.000	0.999	0.984	
(b) S&P 500 Earnings Growth						
Horizon h (quarters)	4	LTG 4-5	LTG 1-10			
Survey/benchmark	IBES	IBES	IBES			
$\text{MSE}_{\mathbb{E}}/\text{MSE}_{\mathbb{F}}$ (LSTM)	0.366	0.750	0.162			
OOS R^2	0.634	0.250	0.838			
w^*	1.000	0.999	1.000			
$\text{MSE}_{\mathbb{F}}/\text{MSE}_{\mathbb{H}}$ (Mean)	1.599	1.143	3.860			

Notes: $\text{MSE}_{\mathbb{E}}$, $\text{MSE}_{\mathbb{F}}$, and $\text{MSE}_{\mathbb{H}}$ are the mean-squared-forecast-errors of machine, benchmark, and historical mean forecasts, respectively. OOS R^2 is defined as 1-$\text{MSE}_{\mathbb{E}}/\text{MSE}_{\mathbb{F}}$. The weight on the machine forecast of for the hybrid forecast is denoted by w^*. Stock returns denote either the CRSP stock market return or S&P 500 price growth, depending on the survey comparison benchmark that most closely aligns with the concept that survey respondents are asked to predict. When the benchmark is the CB, SOC, Gallup/UBS, and CFO survey, we interpret the survey to be asking about log stock returns and compare its forecast against machine forecasts of log CRSP stock returns. When the benchmark is Livingston, we interpret the survey to be asking about log price growth and compare its forecast against machine forecasts of log S&P 500 price growth. Filtered Investor (FI) expectations series aggregates forecasts from Gallup/UBS, CFO, Livingston, and SOC surveys into a latent one-year-ahead expected return using a state-space model. The testing sub-sample is 2005:Q1 to 2023:Q4 for stock returns when the benchmark is CB, SOC, CFO, and FI, 2005:Q1 to 2008:Q4 when the benchmark is the Gallup/UBS survey, and semi-annual over 2005:Q1 to 2023:Q4 for Q2 and Q4 of each calendar year when the benchmark is the Livingston survey. For earnings growth, the testing sub-sample is 2005:Q1 to 2023:Q4, with h-quarter forecasts constructed from the middle month of each quarter. LTG 4-5 and LTG 1-10 denote Long-Term Growth (LTG) forecasts that are interpreted as earnings growth for 4-to-5 and 1-to-10 years, respectively. Forecasts of the historical mean ("Mean") are based on a first estimate of the mean CRSP stock market return from 1959:Q1 to 2004:Q4 and then recursively updating the estimate, adding one observation at a time, over period 2005:Q1 to 2023:Q4.

longer horizon earnings growth rates, finding that LTG exhibited the lowest forecast error when it was used to predict the five-year forward growth rate. The second interpretation follows the implicit assumption in Nagel and Xu (2022).

Panel (b) shows that the machine substantially improves on the predictive accuracy of the IBES analysts, where the ratio of the machine-MSE to IBES-MSE is 0.37 for one-year-ahead earnings growth, implying the machine is 63% more accurate than the median analyst for these predictions. The machine is about 25% more accurate than LTG when forecasting annual earnings growth five years forward and 84% more accurate when forecasting nine-year earnings growth one year forward. This confirms that LTG is a much stronger predictor of 5-year forward earnings growth than annualized 9-year growth rate.

We can also ask how the machine and survey forecasts differ taking into account uncertainty due to sampling error. A common approach is to test the null hypothesis that two models are equally accurate accounting for sampling error. We argue, however, that such tests are less useful or relevant when the objective is to measure belief distortions in survey point forecasts, as here. The reason is that forecasters must report *a* point estimate, regardless of the amount of sampling error around various model forecasts. In this context, the model that is found to be more accurate is always the better choice, even if a statistical test cannot reject that the predictive accuracy of two models is "the same." Moreover, because such tests merely return a binary answer on whether a null hypothesis is rejected or not, they are silent on the practical quantitative question of by how much one model is more accurate than another.

For these reasons, we follow BLM in using a Bayesian perspective to characterize uncertainty over relative forecast performance, adapting the approach of Amisano and Geweke (2017). 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 always being better. In our context, the approach involves solving for the optimal linear combination of the machine and survey forecasts that minimizes the mean square forecast error over our testing sample. We refer to this linear combination as the optimal "hybrid" forecast. The optimal weight placed on the machine forecast, which we denote w^*, is defined as the one that minimizes the hybrid MSE over our testing sample. (See BLM for details.) The weights w^* are reported in Table 1 for each paired forecast comparison. To interpret these numbers, note that if the machine were always better than the survey, w^* would be 1. This happens with most of the forecast comparisons reported in Table 1. If instead the machine were only marginally better than the survey, w^* would be close to 0.5, indicating one would like to place equal weight on the machine and survey forecasts. In all cases, the numbers reported in Table 1 are close to or equal to unity, implying 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.

This gap in forecast accuracy can be viewed a different way, by comparing correlations between the machine's out-of-sample forecasts and subsequent realized returns with the same correlations for the survey forecasts. Table 2 reports these correlations for the machine over the testing sample and for the surveys over subsamples that depend on the data availability of the surveys. Panel A of this table shows that four of our six survey forecasts (inclusive of FI) are negatively correlated with the subsequent realized returns over the period being forecast. The two exceptions are the Livingston and Conference Board surveys, whose forecasts have correlations of 0.18 and 0.13, respectively, with realized future returns. These findings largely confirm those in the literature on subjective beliefs that predate this paper Greenwood and Shleifer (2014). By comparison, the machine forecasts have a correlation of 0.56 with realized future returns. Panel B shows a result that is new to the literature, namely that the correlation between a survey's predictable mistakes, quantified by $Bias_{s,t}[y_{t+h}] \equiv \mathbb{F}_{s,t}[y_{t+h}] - \mathbb{E}_t[y_{t+h}]$, and subsequent returns is much more negative than the correlation between the surveys themselves and subsequent returns. In this case, $Bias_{s,t}$ is negatively correlated with the subsequent realized returns for all surveys, with correlations that range from a high of -0.37 for the Livingston survey to a low of -0.70 for the Gallup/UBS survey. Perhaps unsurprisingly, this shows that once we isolate the prestakes in investor beliefs and distinguish them from ex post forecast error, we find much larger inaccuracies. Interestingly, however, subjective beliefs are most inaccurate when they are excessively optimistic rather than excessively pessimistic.

These results suggest that survey forecast accuracy varies substantially over time depending on how much of the error in subjective beliefs is attributable to predictable mistakes. Figure 2 confirms that this is the case, by showing the period-by-period break-down

Table 2: Correlations of Stock Return Forecasts and Bias with Realized Returns

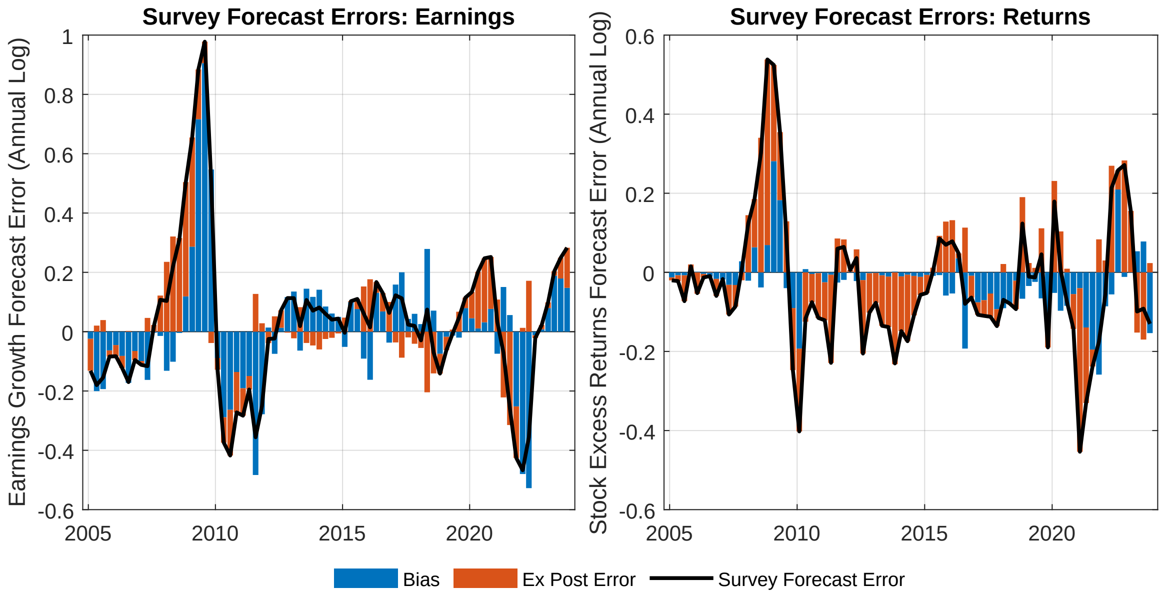
(A) Raw return forecasts vs. realized returns	
Forecast	Correlation
Filtered Investor	-0.190
CFO	-0.191
Livingston	0.179
Gallup/UBS	-0.650
SOC	-0.186
CB	0.132
Machine	0.559
(B) Raw return bias vs. realized raw returns	
Forecast	Correlation
Filtered Investor	-0.574
CFO	-0.589
Livingston	-0.371
Gallup/UBS	-0.698
SOC	-0.577
CB	-0.549

Notes: Reported values are Pearson correlations. Panel A shows correlations between each forecast of the four-quarter ahead raw CRSP value-weighted market return and the corresponding realized raw return. Panel B shows correlations between the bias in each forecast of the four-quarter ahead CRSP value-weighted market return and the corresponding realized raw return. The sample spans 2005:Q1 to 2023:Q4 when the benchmark is CB, SOC, CFO, and FI, 2005:Q1 to 2008:Q4 when the benchmark is the Gallup/UBS survey, and semi-annual over 2005:Q1 to 2023:Q4 for Q2 and Q4 of each calendar year when the benchmark is the Livingston survey.

of survey forecast error, $\mathbb{F}_{s,t}[y_{t+h}] - y_{t+h}$, into the sum of the period-specific prestake, or $Bias_{s,t}[y_{t+h}] \equiv \mathbb{F}_{s,t}[y_{t+h}] - \mathbb{E}_t[y_{t+h}]$, and the ex post error, which is the same as the machine forecast error, $\mathbb{E}_t[y_{t+h}] - y_{t+h}$. The results for IBES expectations of earnings growth are shown on the left, while those for the FI expectations of returns are shown on the right. Figure 2 shows that, in many periods, virtually all of the survey forecast error is attributable to prestakes, although the amount varies substantially over time. For four-quarter-ahead earnings growth expectations, the second half of the GFC and much of its aftermath through 2015 were periods in which IBES forecast errors were driven almost entirely by predictable mistakes. The return survey forecast errors were largely attributable to predictable mistakes in the second half of the GFC, in the period spanning roughly 2016 to 2020, and in the bear market of 2022.

On the other hand, the estimates also show times in which the survey forecast errors were

Figure 2: Sources of Variation in Survey Forecast Errors



Notes: The figure plots survey forecast errors $\mathbb{F}_{s,t}[y_{t+h}] - y_{t+h}$, and decomposes them into the sum of $Bias_{s,t}$ and the ex post error: $\mathbb{F}_{s,t}[y_{t+h}] - y_{t+h} = (\mathbb{F}_{s,t}[y_{t+h}] - \mathbb{E}_t[y_{t+h}]) + (\mathbb{E}_t[y_{t+h}] - y_{t+h})$, where $\mathbb{E}_t[y_{t+h}]$ is the machine forecast. Survey forecasts are from IBES (earnings) and FI (returns). The sample spans 2005Q1-2021Q4.

largely attributable to ex post error, as was the case for the return surveys from roughly 2012 to 2015. In this episode, both the machine and the survey expectations of returns closely tracked the local mean, forecasts that turned out to be too low ex post. Despite this string of ex post mistakes, the estimates imply that the survey forecasts for this period were effectively rational, implying that the high realized returns were largely attributable to a sequence of unpredictable shocks rather than a systematic bias in expectations. Finally, we observe some periods in which a predictable mistake actually reduces the absolute survey forecast error, implying that, in these periods, the machine forecast error exceeded the survey forecast error. These are instances in which the machine uncovered a predictable mistake and yet under-performed by chance. They represent instances of ex post bad luck for objective beliefs rather than ex ante bad judgment.

3.2 Investigating the Prediction Gap with Forecast Decompositions

To better understand the large gap in prediction outcomes between the machine and surveys, we ask the machine to learn about the forecasting models that best explain the survey responses at each point in time and compare them to the specifications for objective beliefs. With these two estimated models in hand, we decompose each into coarse categories of input features, or "themes," that we describe now. In what follows, we let $\mathcal{X}_t$ denote the vector of inputs available as of time t, and χ^t denote the history of inputs from $1, ..., t$.

First consider our measure of objective beliefs, the machine forecasts. The recurrent neural network function estimated at each t can be arbitrarily well approximated as the sum of three components:

$$\mathbb{E}(\cdot) = G(\cdot) \approx \underbrace{b_t}_{\text{local mean}} + \underbrace{h_t(h_{t-1}, \bar{\mathcal{X}}_t; \vec{b}_t, \widehat{\boldsymbol{\beta}}_t)}_{\text{nonlinear history}} + \underbrace{\nabla G(\cdot)' \mathcal{X}_t}_{\text{current information}} . \tag{2}$$

The first, b_t, component is the time t estimate of the outer bias term of the LSTM estimator. This intercept term evolves slowly over time and is akin to a local mean. The second, $h_t(\cdot)$, term is a nonlinear function of the entire history of empirical inputs through $t-1$, holding fixed current inputs. This can be understood by noting that $h_t(\cdot)$ is a function of h_{t-1}, which is a function of $\mathcal{X}^{t-1}$. The $h_t(\cdot)$ function is evaluated at the current estimated vector of LSTM inner bias terms, $\vec{b}_t$, and weights, $\widehat{\boldsymbol{\beta}}_t$, holding fixed current inputs $\mathcal{X}_t$ at their sample mean over the training sample, which is indicated by writing the current input argument as$\bar{\mathcal{X}}_t$. The third, $\nabla G(\cdot)' \mathcal{X}_t$, term captures the impact of current inputs $\mathcal{X}_t$. This is accomplished using a linear approximation that holds fixed the history $\mathcal{X}^{t-1}$. Here, $\nabla G(\cdot)$ denotes the integrated gradient of the LSTM function with respect to $\mathcal{X}_t$. The approximation accounts for non-linearities over $\mathcal{X}_t$ via the gradient magnitudes and can be made arbitrarily accurate by choosing vanishingly small step sizes to approximate the integral. See Appendix section 3.4 for details. This current information component is further decomposed into the broad themes of "macro fundamentals," "non-fundamentals," and "media-sentiment" as discussed above.

The above approximation is used for decomposing machine forecasts over time. For decomposing survey respondent forecasts, we first use a machine algorithm to learn about the forecasting models of the survey respondents. Specifically, we ask an LSTM machine algorithm to predict the survey respondent's forecast out-of-sample, by minimizing the distance between the survey response at time t and the target variable at time t, using the history of input data through $t-1$. To mitigate over-fitting, we estimate recursive out-of-sample forecasts where the loss function is penalized using ridge and lasso penalties. We then decompose the fitted survey forecast models into the same components as the machine forecast, plus a residual:

$$\hat{\mathbb{F}}(\cdot) \approx \underbrace{b_{\mathbb{F},t}}_{\text{local mean}} + \underbrace{h_t(h_{t-1}, \bar{\mathcal{X}}_t; \vec{b}_{\mathbb{F},t}, \widehat{\boldsymbol{\beta}}_{\mathbb{F},t})}_{\text{nonlinear history}} + \underbrace{\nabla G_{\mathbb{F}}(\cdot)'\mathcal{X}_t}_{\text{current information}} + \underbrace{u_t}_{\text{residual}}$$

The residual captures the error in the fitted model and represents input features intangible to the machine, such as private information or forecaster judgment.

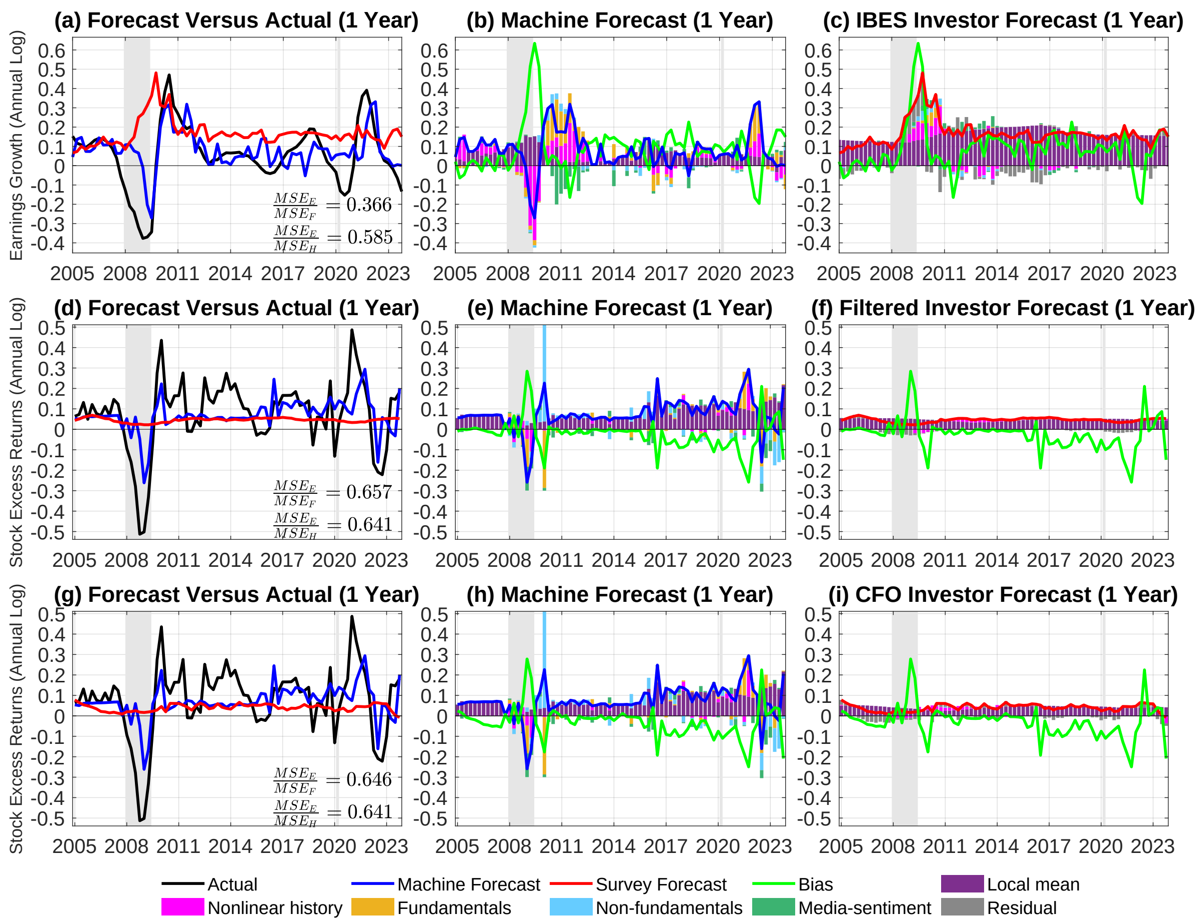
Figure 3 gives a visual impression of the forecasts and their decompositions into broad themes over the testing sample 2005:Q1-2023:Q4. Panels (a), (d), and (g) show machine and survey forecasts of four-quarter-ahead corporate earnings growth or stock market returns along with realized earnings growth or stock market returns during the corresponding four-quarter period being forecast, so that a vertical slice shows the forecast error. The second two columns plot the bias, $Bias_t[y_{t+h}] \equiv \mathbb{F}_t[y_{t+h}] - \mathbb{E}_t[y_{t+h}]$, and also decompose the forecasts into components driven by the local mean, nonlinear history, and current information variables, with the current information variables further decomposed into macro fundamentals, non-fundamentals, and media/sentiment. The bars in these subplots show the contribution of each component to the forecast named in the y axis, where the contributions sum to 100%.[8]

The figure shows that both surveys of earnings growth and returns, but especially subjective return expectations, are extremely stable relative to the volatile realized series. As

[8] Since most inputs are factors that are identified only up to a rotation, it is not possible to infer the sign of the relationship from these plots.

a demonstration that this is not an artifact of our filtering procedure used to obtain the FI return expectation series, the third row shows a similarly stable CFO survey forecast over time. IBES earnings growth expectations are more volatile, but almost never predict negative earnings growth. The third column shows that both the IBES earnings forecasts and

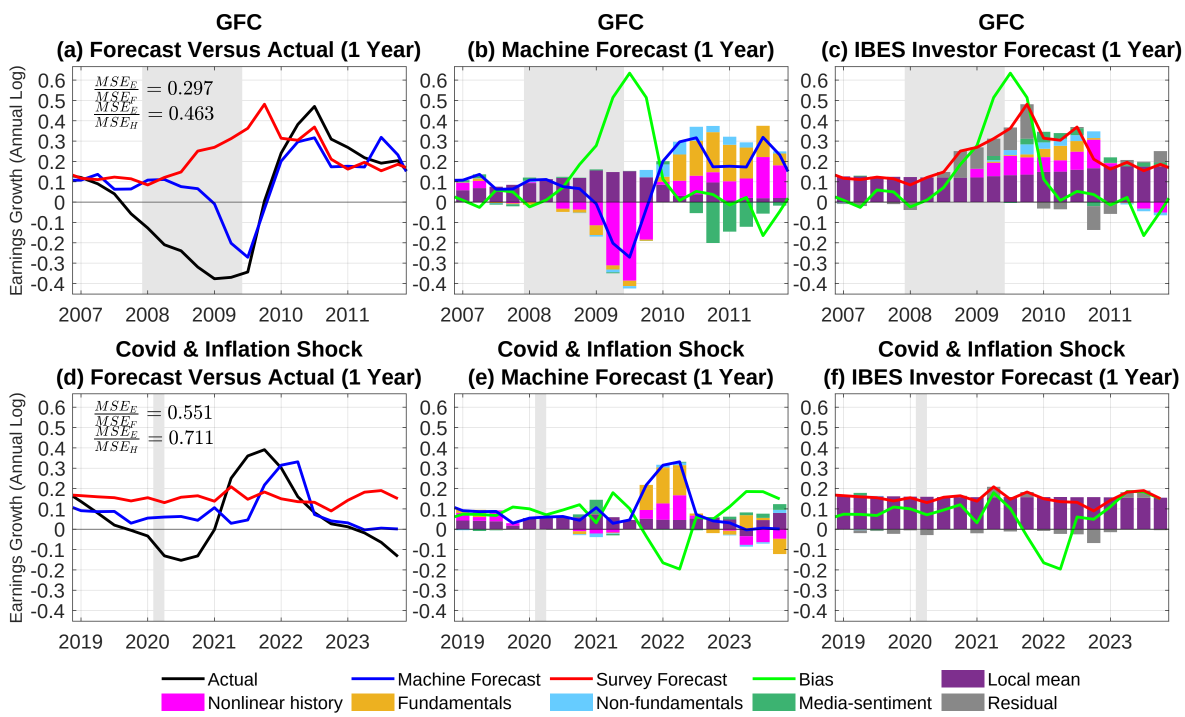
Figure 3: Analysis of Street Earnings Growth and Stock Return Expectations: Full Sample



Notes: The figure plots the machine forecast and median survey forecast of four-quarter-ahead earnings growth or returns over the full testing sample along with actual earnings growth or returns during the corresponding four-quarter period being forecast. $MSE_{\mathbb{E}}$, $MSE_{\mathbb{F}}$, and $MSE_{\mathbb{H}}$ are mean-squared-forecast-errors of the machine forecast, investor forecast, and a recursively estimated historical mean forecast, respectively. The survey forecast for earnings growth is the median IBES equity analyst forecast, while that for stock excess returns is either the Filtered Investor (FI) or median CFO forecast. Stock excess returns are expressed as log annual returns in excess of the 1-year T-bill rate. NBER recessions are shaded in gray. The testing sample spans 2005:Q1-2023:Q4.

the return forecasts closely track the local mean in the target variable much of the time. By contrast, the machine forecast varies more with the nonlinear history, macro fundamentals, and media sentiment, especially during periods of important economic change. The figure also shows a large variation in $Bias_t$ over the testing sample. While it is known that IBES expectations can be biased on average, our interest is in the time-variation in distortions, especially during periods of unusual turbulence. We therefore zoom in on two periods of heightened turbulence in Figures 4 and 5: (i) the GFC and (ii) the period spanning the onset of the Covid-19 pandemic and the subsequent bear market of 2022 that occurred with the spike in general price inflation and lift-off to Fed rate increases.

Figure 4: Analysis of Street Earnings Growth Expectations: Crisis Episodes



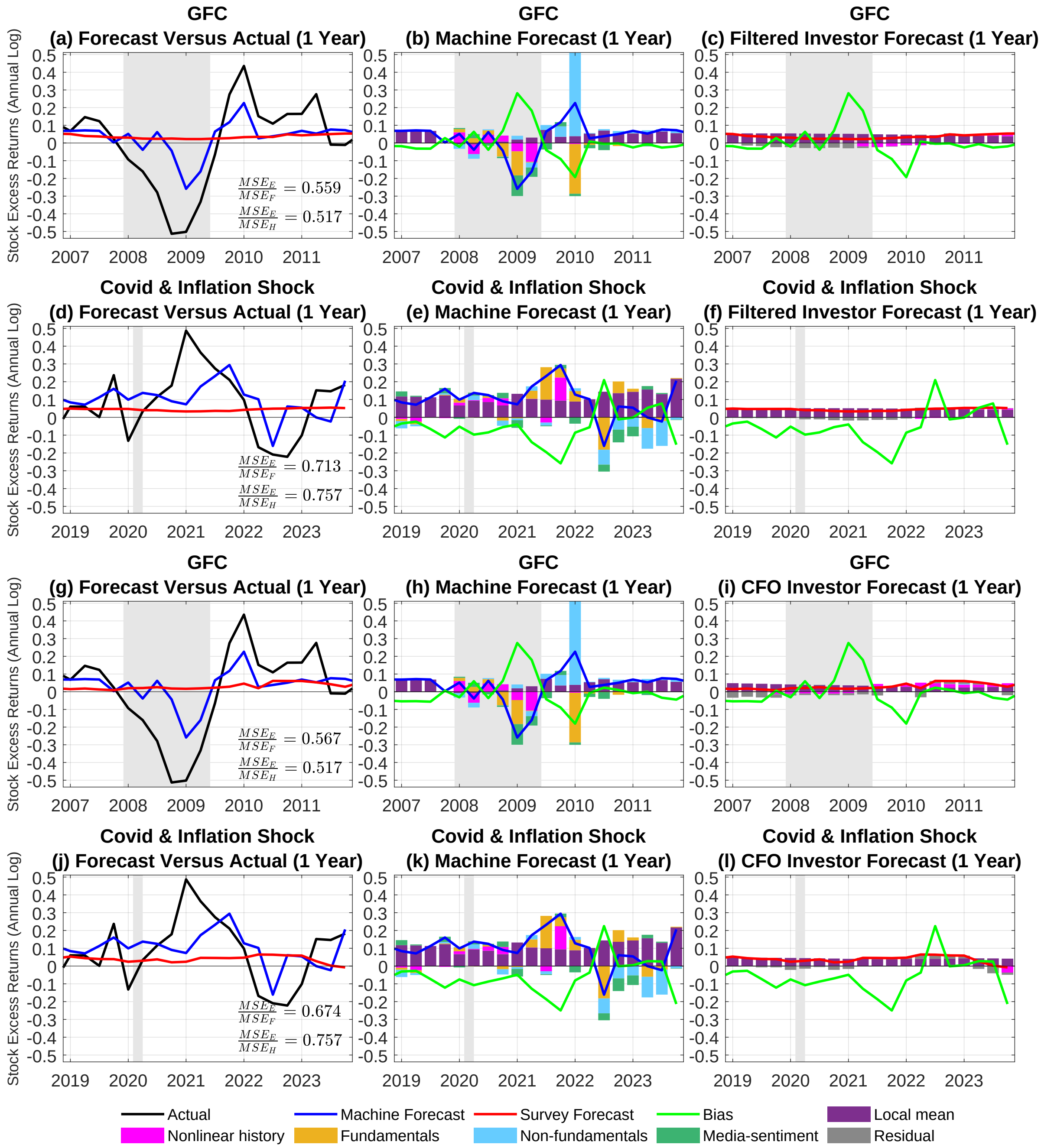
Notes: See Figure 3. The figure plots the observed S&P 500 earnings growth along with the machine forecast, IBES investor forecast, and bias over the Global Financial Crisis (top row) and Covid-19 recession and post-Covid inflation shock (bottom row). The sub-sample plots span 2007:Q1-2011:Q4 (GFC) and 2019:Q1-2023:Q4 (Covid-19 & Inflation Shock).

Several points about these figures bear noting. First, survey respondents have forecasts that are biased up in both the GFC and the bear market of 2022, regardless of whether we consider subjective earnings growth expectations or subjective return expectations. In both episodes, survey expectations continue to be largely explained by the local mean driven by recent trends, a form of recency bias that attaches little weight to long-memory features of the data, reminiscent of the "fading memory" distortion of Nagel and Xu (2022). This contrasts with our benchmark measure of objective beliefs, which places greater weight on the nonlinear history, macro fundamentals, and media sentiment variables especially in crisis episodes in order to identify turning points. Second, macro fundamentals receive relatively little attention from survey respondents of either type, whereas objective beliefs make use of this information especially during or after periods of volatility. Third, Figures 4 and 5 provide little evidence that survey respondents learn from past mistakes. For example, the survey responses fail to predict any meaningful part of these downturns, even with a lag. Objective beliefs on the other hand often miss the early stages of a downturn but course-correct in the middle phase. This is possible because the machine is forced to question its specification by evaluating a range of competing specifications in both the training and validation steps before making a true out-of-sample forecast. During the GFC and 2022 bear market, the machine specification accounts for sharp non-linearities by switching to deeper, wider networks with more lightly regularized penalties for complex specifications, which allows it to predict part of the sharp downturns in those episodes.

To provide more granular detail on what information objective beliefs pick up on in these episodes, Figure 6 zooms in on the most important variables across all input variables and themes. As in the previous figure, the nonlinear history component plays a crucial role in the machine's prediction of negative earnings growth and its sharp subsequent recovery in the GFC, a point we come back to below. For return forecasts, the nadir of the GFC objective belief forecast occurs in 2009:Q1, and is driven mostly by media-sentiment and a macro variable called the "FOMC surprise." The FOMC surprise variable is constructed by accumulating changes in short-dated federal funds futures contract rates from 10 minutes before to 20 minutes after each Federal Reserve Federal Open Market Committee (FOMC) announcement in a given quarter. Figure 6 shows that large cumulative surprises in the fed funds futures market tend to precede market downturns. For the post-Covid period, macro factors related to consumer spending played important roles in predicting the rise and subsequent decline in both objective earnings growth and return forecasts.

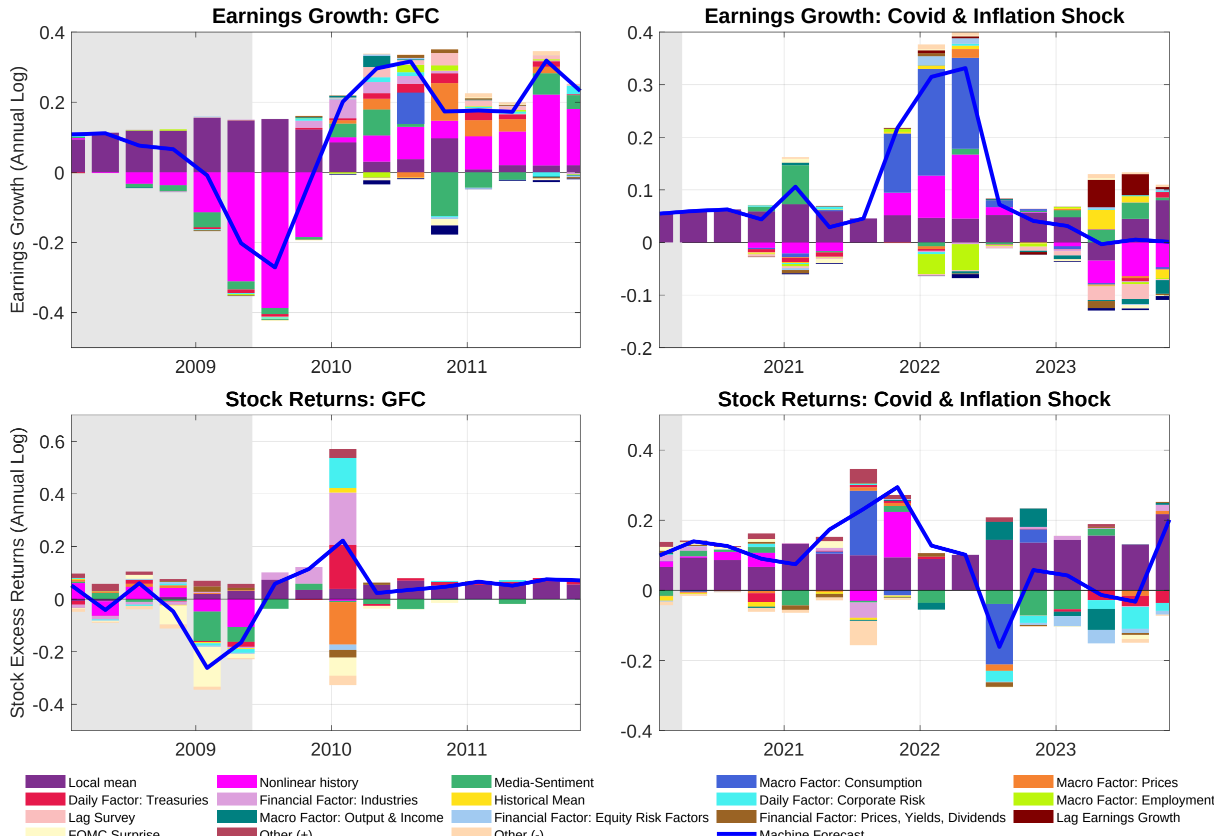
Given the importance of the nonlinear history component in the machine forecasts, we also investigate its core economic drivers. Despite the fact that it is a high-dimensional

Figure 5: Analysis of Stock Return Expectations: Crisis Episodes



Notes: See Figure 3. The figure plots the observed stock excess return along with the machine forecast, investor forecast, and bias over the Global Financial Crisis (top rows) and Covid-19 recession and inflation shock (bottom rows). All series are displayed after subtracting off the 1-year T-bill rate. The survey forecast for stock excess returns is the Filtered Investor (FI) or CFO survey forecast. The sub-sample plots spans: 2007:Q1-2011:Q4 (GFC) and 2019:Q1-2023:Q4 (Covid-19 & Inflation Shock).

Figure 6: Detailed Forecast Decompositions



Notes: The figure plots the machine forecast and its decomposition. The x-axis indicates the forecasted period with realized outcomes. The machine forecast and investor forecast are decomposed into core themes. Taken together, the components represented with colored bars account for 100% of the variation in the machine forecast and investor forecast in each period. For the investor forecast, an additional residual term captures the portion of the forecast not explained by the LSTM model out of sample. Stock excess returns are expressed as log annual returns in excess of the 1-year T-bill rate, and earnings growth is measured as the log annual growth rate. NBER recessions are shaded in gray. The sample spans: 1991:Q1-2004:Q4 (training earnings), 1973:Q1-2004:Q4 (training stock returns), 2005:Q1-2023:Q4 (testing), 2007:Q1-2011:Q4 (GFC) and 2019:Q1-2023:Q4 (Covid-19 & Inflation Shock).

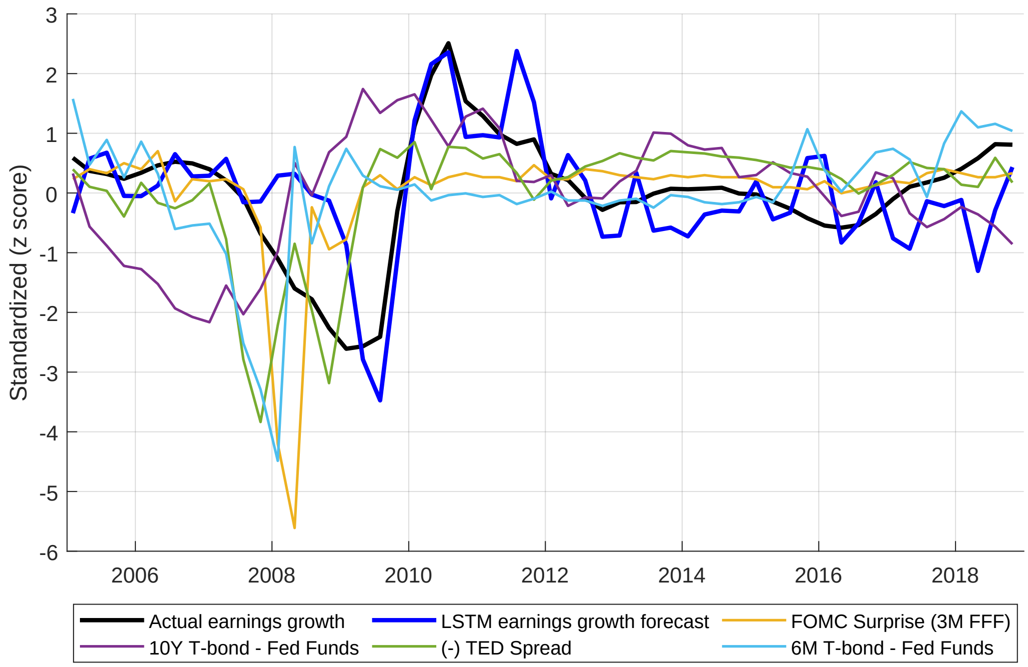
function of the full history of data inputs, we find that much of its variation is driven by a few variables. To name these drivers, Appendix Section 3.4 reports the highest marginal R^2 statistics from regressions of the first or second lags of historical data inputs onto the nonlinear history component. This output shows that the nonlinear history component is highly correlated with variables that proxy for corporate credit risk (such as the TED spread over the portion of our sample when it was available), Treasury market variables, and the previously mentioned FOMC surprise variable. The channeling of this information through the nonlinear history is relevant because this component reflects the LSTM's non-linear hidden layers that move with the entire history of input data through $t-1$, the quarter immediately prior to the current-quarter forecast. This component allows the machine to tap into longer-memory features in recognizing the central roles played by dramatic changes in specific input variables during times of unusual volatility.

A prime example of this is the machine forecast of earnings growth during the GFC, where much of its variation is driven by the nonlinear history term (Figure 4). Figure 7 plots the top empirical contributors to variation in the nonlinear history component during the GFC, along with the machine forecast of earnings growth and realized earnings growth. These top contributors are: the TED spread (plotted as the negative of the TED spread), the 10-year Treasury bond-fed funds rate spread, the 6-month Treasury bill-fed funds rate spread, and the FOMC surprise variable mentioned above. The figure plots the raw series over time, with all series standardized to unit variance for comparability. It is clear that these variables all moved sharply before earnings growth did. The TED spread in particular reached unusually high levels in August of 2007 and by September of 2008 had risen to historically unprecedented levels. Such volatility eventually feeds into the machine's input layer and nonlinear history. By contrast, these variables play little role in the estimated survey forecast specifications, a finding that the analysis concludes was a predictable mistake.

3.3 Subjective vs. Objective Risk-Return Trade-off

Any finding of meaningful predictable mistakes in subjective beliefs raise the possibility that real-world investors have a distorted view of the objective tradeoff between risk and reward. To consider this possibility, we investigate how the risk-return tradeoff in subjective beliefs compares to measures of the same in objective beliefs. For subjective beliefs, we use two measures of subjective risk premia from return surveys, namely our FI series, and the individual investor series used in Nagel and Xu (2022) and Nagel and Xu (2023), which we abbreviate with "NX" below. For measures of subjective risk, we follow Nagel and Xu (2023) in using three measures of perceived risk in the stock market. The first is constructed from

Figure 7: Machine Forecast of Earnings Growth and Top Contributors to Nonlinear History in the GFC



Notes: The figure plots the machine forecast of earnings growth together with inputs that load strongly on the nonlinear history: the negative of the Ted spread, the 10Y Treasury bond-federal funds rate spread (10Y T-bond-Fed Funds), the 6 month Treasury bill-federal funds rate spread (6M T-bill-Fed Funds), and the "FOMC surprise (3M FFF)," equal to the cumulative sum of all changes in the 3 month federal funds futures contract rate from 10 minutes before to 20 minutes after each FOMC announcement in the current quarter up to the date given on the y axis. The TED spread is constructed as the 3-month LIBOR minus the 3-month Treasury bill rate through Jan 2022 and a synthetic LIBOR from Bloomberg thereafter. All series are normalized to unit variance. The x axis marks the period when predicted outcomes were realized.

the CFO survey and provides a quarterly measures of perceived variance over the next 12 months. The second two measure "crash confidence," of either individual or institutional investors, where the measures give the percentage of respondents in each category who attach little probability to a future stock market crash. We take the negative of this index to proxy for perceived crash risk. (See Appendix 3.4 for details.) For the objective beliefs, we use our machine-based measures of objective risk premia along with previously estimated measures of objective financial market uncertainty from Ludvigson, Ma and Ng (2021). Table 3 reports regressions of subjective (objective) expected excess returns on subjective (objective) measures of risk. Since the risk measures are on different scales, we normalize the coefficient by dividing each risk coefficient by the standard deviation of the respective risk measure. Our interest is not in the raw values of the two risk coefficients from these regressions but in their ratio. The table reports the ratio along with Bayesian credible sets for the ratio and the posterior probability that the risk coefficient in subjective beliefs is less than that in objective beliefs, or that the ratio is less than one.

Table 3 shows that we find a positive risk-return tradeoff in both subjective and objective beliefs. However, we find a large difference in magnitudes. Objective risk premia move between 2 and 11.3 times more with objective measures of risk than subjective risk premia move with subjective measures of risk, depending on the subjective measure of risk. These results are similar regardless of whether we use the FI or NX measure of subjective risk premia. Though the credible sets for ratio estimates tend to be wide, as here,[9] the posterior probability that the subjective risk coefficient is less than the objective risk coefficient exceeds 85% in all cases. This suggests that subjective risk premia are sub-optimally responsive to evidence of changing risk, even if we allow for the measures that investors actually perceive.

Panel D of Table 3 juxtaposes these results with those from a related set of regressions. Here we compare the risk coefficient from a regression of subjective risk premia on *objective* rather than subjective risk, with the same coefficient in a regression of objective risk premia on objective risk. This comparison investigates the extent to which subjective risk premia vary sub-optimally with objective risk. We find that the ratio of the former to the latter is much smaller than in the earlier panels, where we compared the full subjective risk-return tradeoff with the objective risk-return tradeoff. Indeed, this ratio is essentially zero when using the NX measure of subjective risk premia. Whether subjective risk premia are measured with NX or FI, the posterior probability subjective risk premia move less than objective risk premia with objective risk exceeds 96% in both cases. What this shows is that subjective risk premia bear virtually no relationship with objective measures of risk, even if they exhibit some relation with subjective measures of risk.

[9]This can happen because some draws from the posterior will include small values for the denominator coefficient.

Table 3: Subjective vs. Objective Risk-Return Trade-off

Survey: $\mathbb{F}^s_t[r^e_{t,t+h}] = \beta_{\mathbb{F},0} + \widetilde{\beta}_{\mathbb{F},1}\widetilde{\mathbb{V}}_t[r_{t,t+h}] + \epsilon_{\mathbb{F},t}$		
Machine: $\mathbb{E}_t[r^e_{t,t+h}] = \beta_{\mathbb{E},0} + \beta_{\mathbb{E},1}\mathbb{V}_t[r_{t,t+h}] + \epsilon_{\mathbb{E},t}$		
Survey s	Filtered Investor (FI)	Nagel and Xu (NX)
Horizon h (quarters)	4	4
(A) Perceived Risk $\widetilde{\mathbb{V}}_t[r_{t,t+h}]$: Crash Confidence Index (Individuals)		
Objective Risk $\mathbb{V}_t[r_{t,t+h}]$: Financial Uncertainty Index		
$\widetilde{\beta}_{\mathbb{F},1}/\beta_{\mathbb{E},1}$	0.255	0.238
90% Credible Set	$[-0.68, 1.41]$	$[-0.70, 1.19]$
$P(\widetilde{\beta}_{\mathbb{F},1}/\beta_{\mathbb{E},1} < 1)$	0.926	0.937
(B) Perceived Risk $\widetilde{\mathbb{V}}_t[r_{t,t+h}]$: Perceived Variance (CFO)		
Objective Risk $\mathbb{V}_t[r_{t,t+h}]$: Financial Uncertainty Index		
$\widetilde{\beta}_{\mathbb{F},1}/\beta_{\mathbb{E},1}$	0.181	0.088
90% Credible Set	$[-0.48, 1.07]$	$[-0.47, 0.75]$
$P(\widetilde{\beta}_{\mathbb{F},1}/\beta_{\mathbb{E},1} < 1)$	0.946	0.964
(C) Perceived Risk $\widetilde{\mathbb{V}}_t[r_{t,t+h}]$: Crash Confidence Index (Institutions)		
Objective Risk $\mathbb{V}_t[r_{t,t+h}]$: Financial Uncertainty Index		
$\widetilde{\beta}_{\mathbb{F},1}/\beta_{\mathbb{E},1}$	0.477	0.366
90% Credible Set	$[-1.43, 2.36]$	$[-1.04, 1.93]$
$P(\widetilde{\beta}_{\mathbb{F},1}/\beta_{\mathbb{E},1} < 1)$	0.850	0.889
(D) Objective Risk $\mathbb{V}_t[r_{t,t+h}]$: Financial Uncertainty Index		
Survey: $\mathbb{F}^s_t[r_{t,t+h}] = \beta_{\mathbb{F},0} + \beta_{\mathbb{F},1}\mathbb{V}_t[r_{t,t+h}] + \epsilon_{\mathbb{F},t}$		
Machine: $\mathbb{E}_t[r_{t,t+h}] = \beta_{\mathbb{E},0} + \beta_{\mathbb{E},1}\mathbb{V}_t[r_{t,t+h}] + \epsilon_{\mathbb{E},t}$		
$\beta_{\mathbb{F},1}/\beta_{\mathbb{E},1}$	0.158	-0.003
90% Credible Set	$[-0.49, 0.91]$	$[-0.58, 0.61]$
$P(\beta_{\mathbb{F},1}/\beta_{\mathbb{E},1} < 1)$	0.955	0.970

Notes: Table reports the ratio of slope coefficients from two regressions: one regressing subjective expectations of stock returns in excess of the prevailing 1-year T-bill rate on perceived risk measures, and the other regressing machine expectations of the same on objective risk measures. We compute the ratio $\widetilde{\beta}_{\mathbb{F},1}/\beta_{\mathbb{E},1}$ using 100,000 posterior draws, where each coefficient is simulated from the marginal t-distribution for the posterior under flat priors (Hamilton, 1994). The 90% credible set is defined by the 5th and 95th percentiles of the posterior distribution. $\mathbb{P}(\widetilde{\beta}_{\mathbb{F},1}/\beta_{\mathbb{E},1} < 1)$ denotes the share of posterior draws where the ratio is less than one. $r^e_{t,t+h}$ denotes the h-quarter–ahead log return on the CRSP value-weighted index (including dividends) in excess of the T-bill. Subjective expectations $\mathbb{F}^s_t[r^e_{t,t+h}]$ are measured from surveys with $s = FI, NX$. All risk coefficients are normalized by dividing by the standard deviation of the corresponding risk measure and multiplying by 100. The sample spans 2005:Q1-2023:Q4.

3.4 Trading Strategies that Bet on Objective beliefs

For our last investigation, we ask whether our measure of objective beliefs could have been exploited to earn positive risk-adjusted returns, or "alpha." Since the machine forecasts can deviate from the survey forecasts only if the algorithm detects predictable mistakes in subjective beliefs, strategies that explicitly bet on objective beliefs are implicitly betting against the predictable mistakes in subjective beliefs.

The trading strategies are indexed by s and differentiated by a prediction-type that now includes the machine forecast or a survey forecast. We denote the prediction of type s at time t as $\mathbb{P}_{s,t}$, where s generically denotes either the machine prediction $\mathbb{E}_t$ or a survey forecast $\mathbb{F}_{s,t}$. We report risk-adjusted returns, or alphas, for strategies that exploit the machine forecasts, i.e., objective beliefs, and compare them to strategies that rely instead on survey forecasts, i.e., subjective beliefs. As an infeasible benchmark, we also report alphas for a perfect foresight strategy in which the CRSP market return was predicted without error. Risk-adjusted returns are estimated relative to several different factor models discussed below as the intercepts from time-series regressions of strategy returns in excess of the prevailing T-bill rate onto factor returns.

We explore two trading strategies that bet on objective beliefs. One is a "long-only" strategy that invests in the CRSP value-weighted market index whenever the machine's forecast of excess returns over the next four quarters is positive and otherwise invests in the 1-year T-bill. To understand this strategy, recall that the machine is trained to make predictions over the next four quarters, with true out-of-sample forecasts made from 2005:Q1 to 2023:Q4. In order to exploit these predictions, portfolio choices starting in 2005 use the four-quarter-ahead machine forecast to decide which one-year investment to make, with the realized return from this decision measured over the following four quarters. Rolling forward, this process is repeated for the next four-quarter period, with re-balancing continuing every four quarters until we reach 2023:Q4. Since there are four ways to do this depending on whether we start in 2005:Q1, 2005:Q2, 2005:Q3 or 2005:Q4, we consider the results from four separate four-quarter-ahead portfolio strategies each year, and report the average alpha from all four. This smooths out sampling noise and mitigates the possibility that the results are overly influenced by an arbitrary starting point. The raw return to this strategy is $r_{s,t,t+4} = r_{t,t+4} \times \mathbf{1}_{\{\mathbb{P}_{s,t}[r_{t,t+4}] > r_{f,t}\}} + r_{f,t} \times \mathbf{1}_{\{\mathbb{P}_{s,t}[r_{t,t+4}] \leq r_{f,t}\}}$, where $\mathbb{P}_{s,t}[r_{t,t+4}]$ is the type-s prediction of market returns over the subsequent four quarters, $r_{t,t+4}$ is the CRSP value-weighted market return over the subsequent four quarters, $r_{f,t}$ is the prevailing 1-year T-bill rate at time t, and $\mathbf{1}_{\{\cdot\}}$ is an indicator function that equals 1 if the statement in curly brackets is true, and zero otherwise.

The second trading strategy is a "long-short" strategy that invests in the CRSP value-weighted market index whenever the excess return forecast is positive, invests in the T-bill if the excess return forecast is negative but the raw return forecast is positive, and invests in the T-bill but short-sells the market if the raw stock return forecast is negative. As before, the four-quarter-ahead forecast is used to decide how to invest, with the realized return measured over the following year. The raw return to this strategy is $r_{s,t,t+4} = r_{t,t+4} \times \mathbf{1}_{\{\mathbb{P}_{s,t}[r_{t,t+4}]>r_{f,t}\}} + r_{f,t} \times \mathbf{1}_{\{0<\mathbb{P}_{s,t}[r_{t,t+4}]\leq r_{f,t}\}} - (r_t - r_{f,t}) \times \mathbf{1}_{\{\mathbb{P}_{s,t}[r_{t,t+4}]\leq 0\}}$. The alphas from these machine-based strategies are compared below to those that would have instead used a survey forecast.

We consider alphas from these strategies relative to two different factor models: the CAPM and the Fama-French 5-factor model (Fama and French (2015)).

Table 4 reports the alphas from the long-only strategy. A strategy that would have bought the CRSP index when the machine excess return forecast was positive and otherwise invested in the T-bill yields earns risk-adjusted returns equal to 4.6 percent per annum relative to the CAPM and 4.4 percent per annum relative to the Fama-French 5-factor model. As a point of comparison, the perfect foresight alphas range between 7 and 8 percent per annum for these two cases. The relatively large machine alphas are the product of defensive strategies, where the alphas are "earned" almost entirely during and after periods of market turbulence such as during the GFC and the bear market of 2022, when the strategies were relatively successful at mitigating losses. By contrast, when trades are made on the basis of survey forecasts, indicated by the entries in the first column labeled FI, CFO, or Livingston forecasts, the alphas are all effectively zero. This happens because the surveys rarely anticipate losses.

Table 5 shows the alphas for the long-short strategy that would have bought the CRSP index when the excess return forecast was positive, and otherwise bought the T-bill while shorting the market when the machine's raw return forecast was negative. When trades are

Table 4: Alphas from Long-Only Strategy

	α	β_m	β_{SMB}	β_{HML}	β_{RMW}	β_{CMA}
(a) CAPM Factor Model $r_{s,t,t+4} - r_{f,t} = \alpha + \beta_m(r_{m,t} - r_{f,t}) + \varepsilon_t$						
Machine	0.046	0.634				
90% Credible Set	[0.031,0.061]	[0.55,0.72]				
FI	-0.000	1.000				
90% Credible Set	[-0.000,-0.000]	[1.00,1.00]				
CFO	-0.002	0.993				
90% Credible Set	[-0.006,0.002]	[0.97,1.01]				
Livingston	0.000	1.000				
90% Credible Set	[0.000,0.000]	[1.00,1.00]				
Perfect Foresight	0.074	0.513				
90% Credible Set	[0.062,0.086]	[0.45,0.58]				
(b) Fama-French 5-Factor Model $r_{s,t,t+4} - r_{f,t} = \alpha + \beta_m(r_{m,t} - r_{f,t}) + \beta_{SMB}SMB_t + \beta_{HML}HML_t + \beta_{RMW}RMW_t + \beta_{CMA}CMA_t + \varepsilon_t$						
Machine	0.043	0.625	-1.071	0.274	0.577	0.794
90% Credible Set	[0.03,0.06]	[0.54,0.71]	[-2.12,-0.03]	[-0.82,1.38]	[-0.71,1.88]	[-0.86,2.44]
FI	0.000	1.000	-0.000	0.000	-0.000	0.000
90% Credible Set	[-0.00,0.00]	[1.00,1.00]	[-0.00,-0.00]	[0.00,0.00]	[0.00,0.00]	[-0.00,0.00]
CFO	0.001	0.989	0.176	-0.087	-0.434	-0.825
90% Credible Set	[-0.00,0.00]	[0.97,1.01]	[-0.07,0.43]	[-0.36,0.17]	[-0.75,-0.12]	[-1.22,-0.43]
Livingston	-0.000	1.000	0.000	-0.000	-0.000	-0.000
90% Credible Set	[-0.00,0.00]	[1.00,1.00]	[-0.00,0.00]	[-0.00,0.00]	[-0.00,0.00]	[-0.00,0.00]
Perfect Foresight	0.072	0.503	-1.203	-0.196	0.601	1.041
90% Credible Set	[0.06,0.08]	[0.44,0.56]	[-1.95,-0.45]	[-1.00,0.60]	[-0.36,1.56]	[-0.17,2.28]

Notes: This table reports regression-based alphas from regressions of trading strategy returns in excess of the 1-year T-bill rate on excess factor returns over the period 2005:Q1-2023:Q4. The trading strategy invests in the CRSP value-weighted market index when an excess return forecast is positive, and otherwise invests in the 1-year T-bill: $r_{s,t,t+4} = r_{t,t+4} \times \mathbf{1}_{\{\mathbb{P}_{s,t}[r_{t,t+4}]-r_{f,t}>0\}} + r_{f,t} \times \mathbf{1}_{\{\mathbb{P}_{s,t}[r_{t,t+4}]-r_{f,t}\leq 0\}}$. Alphas represent the additional annualized log return of the trading strategy after controlling for the specified risk factors. Panel (a) reports CAPM results controlling for market excess returns only. Panel (b) adds size (SMB), value (HML), profitability (RMW), and investment (CMA) factors. 90% credible sets are based on 10,000 posterior draws from the marginal t-distribution under flat priors (Hamilton, 1994), given by the 5th and 95th percentiles of the posterior distribution.

made on the basis of the machine forecast, this strategy would have earned alphas of 9 and 8 percent per annum relative to the CAPM and Fama-French 5-factor models, respectively. This may be compared to the infeasible perfect foresight alphas which range between 14 and 15 percent per annum. When trades are made on the basis of the survey forecasts, alphas are again close to zero. This happens because the surveys rarely anticipate returns that fall below the T-bill rate, let alone losses. Also of interest, the market betas all hover around 0.3 for the betting on objective beliefs strategies. The factor exposure to SMB is negative when evaluated against the Fama-French 5-factor model. This implies that these strategies can earn sizable alphas while simultaneously hedging out some forms of factor risk. On the other hand, factor exposures to the profitability and investment factors in the 5-factor model are substantially greater than one, implying that alphas are earned by trading off defensiveness with respect to some factors with aggressiveness toward others.

While not a definitive test, we view these results as suggestive of a connection between prestakes in subjective beliefs and realized stock market dynamics. After all, if our measured subjective beliefs and their identified distortions were insufficiently representative of "market" beliefs and "market" distortions, we would not expect trading on the basis of the identified distortions to be profitable. Moreover, the greater volatility of spot prices relative to survey forecasts does not preclude the possibility that the subjective measures accurately reflect market beliefs. Both survey respondents and market participants may respond to positive or negative news through intensive buying or selling that moves current prices, even if they never revise their forecasts of *future* returns. From their perspective, large forecast errors for earnings and returns are likely to be interpreted as unpredictable shocks, rather than as partly driven by systematic biases in their own beliefs.

Conclusion

This paper uses a machine learning algorithm to establish a practical measure of rational and efficient expectation formation while identifying distortions and inefficiencies in the subjective beliefs of stock market participants. We extend the existing literature by going beyond the study of subjective beliefs and documenting violations of FIRE, to quantifying their importance in a real-world, real-time context. We find that the overall magnitude of violations is sizable on average, and varies substantially over time. Subjective beliefs often make predictable mistakes, especially in times of market turbulence.

To investigate the nature of these distortions, we train a machine to learn about the forecasting models that best explain the survey responses at each point in time. We find

Table 5: Alphas from Long-Short Strategy

	α	β_m	β_{SMB}	β_{HML}	β_{RMW}	β_{CMA}
(a) CAPM Factor Model $r_{s,t,t+4} - r_{f,t} = \alpha + \beta_m(r_{m,t} - r_{f,t}) + \varepsilon_t$						
Machine	0.087	0.305				
90% Credible Set	[0.058,0.116]	[0.16,0.46]				
FI	-0.000	1.000				
90% Credible Set	[-0.000,-0.000]	[1.00,1.00]				
CFO	-0.002	0.993				
90% Credible Set	[-0.006,0.002]	[0.97,1.01]				
Livingston	0.000	1.000				
90% Credible Set	[0.000,0.000]	[1.00,1.00]				
Perfect Foresight	0.142	0.063				
90% Credible Set	[0.119,0.165]	[-0.06,0.19]				
(b) Fama-French 5-Factor Model $r_{s,t,t+4} - r_{f,t} = \alpha + \beta_m(r_{m,t} - r_{f,t}) + \beta_{SMB}SMB_t + \beta_{HML}HML_t + \beta_{RMW}RMW_t + \beta_{CMA}CMA_t + \varepsilon_t$						
Machine	0.081	0.294	-1.908	0.533	1.367	1.556
90% Credible Set	[0.05,0.11]	[0.13,0.46]	[-3.88,0.08]	[-1.66,2.71]	[-1.16,3.90]	[-1.62,4.66]
FI	0.000	1.000	-0.000	0.000	-0.000	0.000
90% Credible Set	[0.00,0.00]	[1.00,1.00]	[-0.00,-0.00]	[0.00,0.00]	[0.00,0.00]	[-0.00,0.00]
CFO	0.001	0.989	0.176	-0.087	-0.434	-0.825
90% Credible Set	[-0.00,0.00]	[0.97,1.01]	[-0.07,0.42]	[-0.34,0.18]	[-0.75,-0.11]	[-1.22,-0.43]
Livingston	-0.000	1.000	0.000	-0.000	-0.000	-0.000
90% Credible Set	[-0.00,0.00]	[1.00,1.00]	[-0.00,0.00]	[-0.00,0.00]	[-0.00,0.00]	[-0.00,0.00]
Perfect Foresight	0.137	0.046	-2.284	-0.392	1.262	2.054
90% Credible Set	[0.11,0.16]	[-0.08,0.17]	[-3.76,-0.79]	[-1.97,1.22]	[-0.63,3.17]	[-0.34,4.42]

Notes: See Table 4. The trading strategy invests in the CRSP value-weighted market index when the excess return forecast is positive. If the excess return forecast is negative but the raw return forecast is positive, the strategy invests in the T-bill, otherwise short-sells the market and invests in the T-bill: $r_{s,t,t+4} = r_{t,t+4} \times \mathbf{1}_{\{\mathbb{P}_{s,t}[r_{t,t+4}]>r_{f,t}\}} + r_{f,t} \times \mathbf{1}_{\{0<\mathbb{P}_{s,t}[r_{t,t+4}]\leq r_{f,t}\}} - (r_{t,t+4} - r_{f,t}) \times \mathbf{1}_{\{\mathbb{P}_{s,t}[r_{t,t+4}]\leq 0\}}$, where s indexes the forecast $\mathbb{P}_{s,t}[r_{t+4}]$ for market returns over the subsequent 4 quarters. All strategies assume no leverage constraints.

that survey responses exhibit a form recency bias with forecasts that are largely explained by a local mean in the target variable, even when evidence implies that such a prediction has been demonstrably inferior to specifications that utilize longer memory, nonlinear dynamics, and more extensive information sets. We also find that subjective beliefs pay too little attention to early signs of trouble by ignoring volatility in key series such as credit risk indicators, and subsequently display little evidence of learning from past mistakes. These features contrast with our measure of objective beliefs, which adaptively learns from previous episodes of volatility and often course-corrects its specification and forecast to allow for evidence of a sharp turning point. Predictable mistakes imply that subjective risk premia are sub-optimally responsive to evidence of changing risk, even if we allow for the measures that investors actually perceive. Trading strategies that bet against prestakes in subjective beliefs could have earned sizable risk-adjusted returns in the period spanning 2005 to the end of 2023.

References

Adam, K., D. Matveev, and S. Nagel (2021): "Do survey expectations of stock returns reflect risk adjustments?" *Journal of Monetary Economics*, 117, 723–740.

Afrouzi, H., S. Y. Kwon, A. Landier, Y. Ma, and D. Thesmar (2023): "Overreaction in expectations: Evidence and theory," *The Quarterly Journal of Economics*, 138, 1713–1764.

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

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

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

Bastianello, F., P. H Décaire, and M. Guenzel (2024): "Mental models and financial forecasts," *Marius, Mental Models and Financial Forecasts (October 30, 2024)*.

Bauer, M. D., and E. T. Swanson (2023): "An Alternative Explanation for the "Fed Information Effect"," *American Economic Review*, 113, 664–700.

Bianchi, F., S. C. Ludvigson, and S. Ma (2022): "Belief Distortions and macroeconomic fluctuations," *American Economic Review*, 112, 2269–2315.

——— (2024): "What Hundreds of Economic News Events Say About Belief Overreaction in the Stock Market,"Technical report, National Bureau of Economic Research, No. w32301.

Blei, D. M., A. Y. Ng, and M. I. Jordan (2003): "Latent dirichlet allocation," *Journal of machine Learning research*, 3, 993–1022.

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

——— (2022): "Belief Overreaction and Stock Market Puzzles," *Journal of Polticial Economy,* forthcoming.

Bordalo, P., N. Gennaioli, Y. Ma, and A. Shleifer (2020): "Overreaction in macroeconomic expectations," *American Economic Review*, 110, 2748–2782.

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

Bybee, L., B. T. Kelly, A. Manela, and D. Xiu (2021): "Business news and business cycles,"Technical report, National Bureau of Economic Research.

Chen, L., Z. Da, and X. Zhao (2013): "What drives stock price movements?" *The Review of Financial Studies*, 26, 841–876.

Cong, L. W., K. Tang, J. Wang, and Y. Zhang (2021): "Deep sequence modeling: Development and applications in asset pricing," *The Journal of Financial Data Science*, 3, 28–42.

De La O, R., and S. Myers (2021): "Subjective cash flow and discount rate expectations," *The Journal of Finance*, 76, 1339–1387.

——— (2023): "Which Subjective Expectations Explain Asset Prices?", Manuscript, Wharton.

Fama, E. F., and K. R. French (2015): "A five-factor asset pricing model," *Journal of Financial Economics*, 116, 1–22.

Gabaix, X. (2019): "Behavioral Inattention," in *Handbook of Behavioral Economics, Vol. 2* ed. by Bernheim, D., DellaVigna, S., and Laibson, D. New York, NY: Elsivier, 261–343.

Goyal, A., and I. Welch (2008): "A Comprehensive Look at the Empirical Performance of Equity Premium Prediction," *Review of Financial Studies*, 21, 1455–1508.

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

Gu, S., B. Kelly, and D. Xiu (2020): "Empirical asset pricing via machine learning," *The Review of Financial Studies*, 33, 2223–2273.

Hillenbrand, S., and O. McCarthy (2021): "Heterogeneous investors and stock market fluctuations," *Available at SSRN 3944887*.

——— (2024): "Street Earnings: Implications for Asset Pricing," *Available at SSRN 4892475*.

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

Kohlhas, A. N., and A. Walther (2021): "Asymmetric attention," *American Economic Review*, 111, 2879–2925.

Ludvigson, S. C., S. Ma, and S. Ng (2021): "COVID-19 and the Costs of Deadly Disasters," in *AEA Papers and Proceedings* Volume 111, 366–70.

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

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, 1295–1328.

Nagel, S. (2021): *Machine learning in asset pricing*: Princeton University Press.

Nagel, S., and Z. Xu (2022): "Asset pricing with fading memory," *The Review of Financial Studies*, 35, 2190–2245.

——— (2023): "Dynamics of subjective risk premia,"Technical Report 2.

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

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 Aghion, P., Frydman, R., Stiglitz, J., and Woodford, M. Cambridge MA: Princeton University Press, 25–58.

Online Appendix

Additional Results

Correlations of Subjective v.s. Objective Forecasts with Subsequent Realized Returns

This section reports the correlation between the machine or survey forecast, and the subsequent realized excess return over the period being forecast, over the testing sample 2005:Q1-2023:Q4. The first table in this section reports the correlations for raw returns; the second for excess returns.

Table A.1: Correlations of Stock Excess Return Forecasts with Realized Stock Excess Returns

Forecast	Correlation
Filtered Investor	0.208
CFO	0.097
Livingston	0.270
Gallup/UBS	-0.017
SOC	0.193
CB	0.331
Machine	0.595

Notes: Reported values are Pearson correlations between each forecast of the four-quarter ahead CRSP value-weighted market return in excess of the three-month Treasury bill rate and the corresponding realized excess return. The testing sub-sample is 2005:Q1 to 2023:Q4 when the forecast is either the machine, CB, SOC, CFO, and FI, 2005:Q1 to 2008:Q4 when the forecast is the Gallup/UBS survey, and semi-annual over 2005:Q1 to 2023:Q4 for Q2 and Q4 of each calendar year when the forecast is the Livingston survey.

Bayesian Credible Sets

We compute Bayesian credible sets for the ratio of two regression slope coefficients, $\theta = \beta_{\mathbb{F},1}/\beta_{\mathbb{E},1}$, based on its posterior distribution. The two regressions are:

$$\text{Subjective Risk-Return Tradeoff:} \quad y_t^{\mathbb{F}} = X_t^{\mathbb{F}\prime}\beta_{\mathbb{F}} + \epsilon_t^{\mathbb{F}}, \quad \epsilon_t^{\mathbb{F}} \sim \mathcal{N}(0, \sigma_{\mathbb{F}}^2)$$
$$\text{Objective Risk-Return Tradeoff:} \quad y_t^{\mathbb{E}} = X_t^{\mathbb{E}\prime}\beta_{\mathbb{E}} + \epsilon_t^{\mathbb{E}}, \quad \epsilon_t^{\mathbb{E}} \sim \mathcal{N}(0, \sigma_{\mathbb{E}}^2)$$

In the subjective risk-return tradeoff regression, $y_t^{\mathbb{F}}$ represents survey-based return expectations, denoted as $\mathbb{F}_t^s[r_{t,t+h}]$, while in the objective risk-return tradeoff, $y_t^{\mathbb{E}}$ represents machine-based return expectations, denoted as $\mathbb{E}_t[r_{t,t+h}]$. The vector of regressors for the subjective regression is given by $X_t^{\mathbb{F}} = [1, \widetilde{\mathbb{V}}_t[r_{t,t+h}]]'$, which includes an intercept term and subjective risk. For the objective regression, the regressor vector is $X_t^{\mathbb{E}} = [1, \mathbb{V}_t[r_{t,t+h}]]'$, including an intercept term and objective risk. The parameter vectors for the two regressions are $\beta_{\mathbb{F}} = [\beta_{\mathbb{F},0}, \beta_{\mathbb{F},1}]'$ for the subjective regression and $\beta_{\mathbb{E}} = [\beta_{\mathbb{E},0}, \beta_{\mathbb{E},1}]'$ for the objective regression.

Posterior Under Flat Priors Under flat priors for both the regression coefficients β and the residual variance σ^2, the marginal posterior density of the k regression coefficients in a linear regression is a k-dimensional t-distribution with T degrees of freedom, mean $\hat{\beta}$, and scale matrix $\hat{\sigma}^2(X'X)^{-1}$, where T is the sample size, $\hat{\beta}$ are the OLS estimates, and $\hat{\sigma}^2$ is the maximum likelihood estimate of the residual variance. See Hamilton (1994), *Time Series Analysis*, pages 354–357, result (b) of Proposition 12.3, for the derivation of this result under flat priors.[1] Specifically:

$$\beta_{\mathbb{F}} \mid y^{\mathbb{F}}, X^{\mathbb{F}} \sim \mathcal{T}_{T_{\mathbb{F}}}\left(\hat{\beta}_{\mathbb{F}}, \hat{\sigma}_{\mathbb{F}}^2 (X^{\mathbb{F}\prime} X^{\mathbb{F}})^{-1}\right)$$
$$\beta_{\mathbb{E}} \mid y^{\mathbb{E}}, X^{\mathbb{E}} \sim \mathcal{T}_{T_{\mathbb{E}}}\left(\hat{\beta}_{\mathbb{E}}, \hat{\sigma}_{\mathbb{E}}^2 (X^{\mathbb{E}\prime} X^{\mathbb{E}})^{-1}\right)$$

where $\mathcal{T}_T(\mu, \Sigma)$ denotes a k-dimensional t-distribution with T degrees of freedom, mean vector μ, and scale matrix Σ. The OLS estimator for the coefficients is $\hat{\beta} = (X'X)^{-1}X'y$, and the estimated residual variance is computed as $\hat{\sigma}^2 = \frac{1}{T-k}\sum_{t=1}^{T}(y_t - X_t\hat{\beta})^2$, with $k = 2$ being the number of parameters in each regression.

[1] This approach accounts for both parameter estimation uncertainty and residual variance uncertainty in a fully Bayesian framework, without requiring separate draws for σ^2.

Simulation Procedure For $m = 1, \dots, M$ posterior draws, with $M = 100{,}000$ for our baseline results, the simulation proceeds as follows:

1. Draw a realization of the subjective regression coefficients $\beta_{\mathbb{F}}^{(m)}$ from the multivariate t-distribution $\mathcal{T}_{T_{\mathbb{F}}}(\hat{\beta}_{\mathbb{F}}, \hat{\sigma}_{\mathbb{F}}^2(X^{\mathbb{F}\prime}X^{\mathbb{F}})^{-1})$.

2. Draw a realization of the objective regression coefficients $\beta_{\mathbb{E}}^{(m)}$ from the multivariate t-distribution $\mathcal{T}_{T_{\mathbb{E}}}(\hat{\beta}_{\mathbb{E}}, \hat{\sigma}_{\mathbb{E}}^2(X^{\mathbb{E}\prime}X^{\mathbb{E}})^{-1})$.

3. For each draw, compute the ratio of interest: $\theta^{(m)} = \frac{\beta_{\mathbb{F},1}^{(m)}}{\beta_{\mathbb{E},1}^{(m)}}$

Based on the simulated posterior distribution for θ, we report two quantities of interest. First, we construct the 90% Bayesian credible set for θ, defined as the interval between the 5th and 95th percentiles of the simulated draws $\theta^{(m)}$. Second, we compute the posterior probability $\mathbb{P}(\theta < 1)$, given by the fraction of simulated draws for which $\theta^{(m)} < 1$. This provides a Bayesian measure of evidence against the hypothesis that subjective risk-return sensitivity is at least as strong as objective sensitivity.

Data

S&P Market Cap, Indexes, S&P Futures, S&P Dividends, Stock Market Returns, and Treasury Bill Data

We use data on both stock market returns (price growth plus a dividend yield) and on stock market price growth. Monthly data on stock returns are obtained from the Center for Research in Security Prices (CRSP), downloaded from WRDS: `https://wrds.wharton.upenn.edu/wrdsauth/members.cgi`. We use the CRSP value-weighted monthly return series `VWRETD` (which includes dividends) and compute annualized log returns as $\text{lnCRSPD} = 12\ln(1 + \text{VWRETD})$. To measure excess returns we take the difference between the return and the known, i.e., lagged log of the 1-year T-bill rate (1YTB). Since 1YTB is reported at an annual rate in percent, we compute the annualized (raw unit) log of future returns less the current short rate: $\ln(\text{VWREX}_{t+12}) \equiv 12\ln(1 + \text{VWRETD}_{t+12}) - \ln\left(1 + \frac{1\text{YTB}_t}{100}\right)$. Both series were downloaded from WRDS on July 30, 2025. When evaluating the MSE ratios of the machine relative to a benchmark survey, we compute machine forecasts for either the annual CRSP return or S&P 500 price growth, depending on which aligns more closely with the concept survey respondents are asked to predict. To measure one-year stock market price growth, we use the one-year log cumulative growth rate of the S&P 500 index, $\log\left(\frac{P_{t+12}^{S\&P}}{P_t^{S\&P}}\right)$.

The monthly S&P 500 index series spans 1957:03 to 2023:12 and was downloaded from WRDS on July 30, 2025, from the Annual Update of the S&P 500 Index File. To compute the one-year log CRSP return, we use: $\sum_{j=1}^{12} \ln(1 + \text{VWRETD}_{t+j}) - \ln\left(1 + \frac{1\text{YTB}_t}{100}\right)$.

Earnings Data

IBES ("Street") Earnings Data We use IBES "street earnings" earnings per share (EPS) data that start in 1983:Q4 as the forecast target for IBES analysts. Following the recommendation of Hillenbrand and McCarthy (2024), we use Street earnings as the forecast target for IBES analysts. Street earnings differ from GAAP earnings by excluding discontinued operations, extraordinary charges, and other non-operating items. According to the IBES user guide, analysts submit forecasts after backing out these transitory components, and IBES constructs the realized series to align with those forecasts. While analysts have some discretion over which items to exclude, Hillenbrand and McCarthy (2024) demonstrate that the target of these forecasts corresponds closely to earnings before special items in Compustat, suggesting that street earnings accurately reflect the measure analysts are targeting. To aggregate firm-level street earnings to the S&P 500 level, we follow the same methodology used by De La O and Myers (2021) and Hillenbrand and McCarthy (2024) for constructing aggregate earnings measures:

$$E_t = \Omega_t \left[\sum_{i \in x_t} EPS_{i,t} S_{i,t} \right] / Divisor_t$$

where E_t is aggregate S&P 500 earnings, $EPS_{i,t}$ is the street earnings per share of firm i among all S&P 500 firms x_t for which we have data in IBES, $S_{i,t}$ is shares outstanding of firm i, and $Divisor_t$ is the S&P 500 divisor available at: `https://ycharts.com/indicators/sp_500_divisor`. We multiply this aggregate by Ω_t, a ratio of total S&P 500 market value to the market value of the companies with available data to account for the fact that IBES does not provide earnings data for all firms in the S&P 500 in every period. The final quarterly total earnings series spans the period 1983:Q4 to 2023:Q4. We downloaded IBES street earnings data from WRDS on July 19, 2025. The divisor data were downloaded on July 21, 2025.

Survey Data on Earnings Expectations

IBES Survey We obtained the monthly survey data for the median analyst earnings per share forecast and actual earnings per share from the Institutional Brokers Estimate System (IBES) via Wharton Research Data Services (WRDS). The data spans the period 1976:01 to 2023:12.

We build bottom-up measures of aggregate S&P 500 earnings expectations growth using the constituents of the S&P 500 at each point in time. IBES calculates consensus estimate (the median or mean) for any firm only if there are a sufficient number of individual forecasts. This means that any company included in the aggregated S&P 500 forecast has met a minimum threshold of analyst interest. Moreover, reports from FactSet and S&P Global consistently imply that the number of companies with a consensus estimate is essentially the entire 500-member universe, with very rare exceptions. This ensures that the bottom-up forecast is a reliable and comprehensive aggregation of individual firm expectations.

We first construct expected earnings expectations for aggregate earnings h-months-ahead as

$$\mathbb{F}_t[E_{t+h}] = \varOmega_t \left[\sum_{i \in x_{t+h}} \mathbb{F}_{med,t}\left[EPS_{i,t+h}\right] S_{i,t} \right] / Divisor_t,$$

where E is aggregate S&P 500 earnings, $\mathbb{F}_{med,t}\left[EPS_{i,t+h}\right]$ is the median analyst forecast of earnings per share (EPS) for firm i, EPS_i is earning per share of firm i among all S&P 500 firms x_{t+h} for which we have forecasts in IBES for $t+h$, S_i is shares outstanding of firm i, and $Divisor_t$ is calculated as the S&P 500 market capitalization divided by the S&P 500 index. We obtain the number of outstanding shares for all companies in the S&P500 from Compustat. (All data from Compustat were downloaded on July 19, 2025.) IBES estimates are available for most but not all S&P 500 companies. Following De La O and Myers (2021), we multiply this aggregate by $\varOmega_{t+h}$, a ratio of total S&P 500 market value to the market value of the forecasted companies at $t+h$ to account for the fact that IBES does not provide earnings forecasts for all firms in the S&P 500 in every period.

IBES database contains median earnings forecasts up to five annual fiscal periods (FY1 to FY5) and as a result, we interpolate across the different horizons to obtain the expectation over the next X months, as needed. This procedure has been used in the literature, including De La O and Myers (2021). For example, if we are interested in the expectation over the next 12 months, and if the fiscal year of firm XYZ ends nine months after the survey date, we have a 9-month earning forecast $\mathbb{F}_t[E_{t+9}]$ from FY1 and a 21-month forecast $\mathbb{F}_t[E_{t+21}]$ from FY2. We then obtain the 12-month ahead forecast by interpolating these two forecasts as follows,

$$\mathbb{F}_t[E_{t+12}] = \frac{9}{12}\mathbb{F}_t[E_{t+9}] + \frac{3}{12}\mathbb{F}_t[E_{t+21}].$$

where the inputs to this interpolation, $\mathbb{F}_t[E_{t+9}]$ and $\mathbb{F}_t[E_{t+21}]$, are the aggregate S&P 500 earnings expectations for those horizons, constructed from the firm-level median forecasts as detailed above.

For the forecasting performance estimates, we use quarterly data. To convert the monthly forecast to quarterly frequency, we use the forecast made in the middle month of each quarter, and construct one-year earnings expectations from 1976:Q1 to 2023:Q4 and the earning expectation growth is calculated as an approximation following following De La O and Myers (2021):

$$\mathbb{F}_t\left(\Delta e_{t+12}\right) \approx \ln\left(\mathbb{F}_t[E_{t+12}]\right) - e_t$$

where e_t is log earnings for S&P 500 at time t calculated as $e_t = \ln\left(EPS_t \cdot Divisor_t\right)$, where EPS_t is the IBES street earnings per share for the S&P 500, as described above.

We constructed long term expected earnings growth (LTG) for the S&P 500 following Bordalo et al. (2019). Specifically, we obtained the median firm-level LTG forecast from IBES, and aggregate the value-weighted firm-level forecasts,

$$LTG_t = \sum_{i=1}^{S} LTG_{i,t} \frac{P_{i,t}Q_{i,t}}{\sum_{i=1}^{S} P_{i,t}Q_{i,t}}$$

where S is the number of firms in the S&P 500 index, and where $P_{i,t}$ and $Q_{i,t}$ are the stock price and the number of shares outstanding of firm i at time t, respectively. $LTG_{i,t}$ is the median forecast of firm i's long term expected earnings growth. The data spans the periods from 1981:12 to 2023:12. All data were downloaded in July 19, 2025.

To estimate any biases in IBES analyst forecasts, our dynamic machine algorithm takes as an input a likely date corresponding to information analysts could have known at the time of their forecast. IBES does not provide an explicit deadline for their forecasts to be returned. Therefore we instead use the "statistical period" day (the day when the set of summary statistics was calculated) as a proxy for the deadline. We set the machine deadline to be the day before this date. The statistical period date is typically between day 14 and day 20 of a given month, implying that the machine deadline varies from month to month. As the machine learning algorithm uses mixed-frequency techniques adapted to quarterly sampling intervals, while the IBES forecasts are monthly, we compare machine and IBES analyst forecasts as of the middle month of each quarter, considering 12-month ahead forecast from the beginning of the month following the survey month.

Survey Data on Stock Return Expectations

This section describes survey data on stock return expectations. We use several surveys to measure stock return expectations. Some ask about price growth expectations. We adjust these with the dividend yield to arrive at a subjective return expectation. For each subjective one-year ahead expected stock return series, we construct the subjective risk premium by subtracting the (annualized) 1-year Treasury yield that prevailed at the time of the survey.

Gallup/UBS Survey The UBS/Gallup is a monthly survey of one-year-ahead stock market return expectations, obtained from Roper iPoll: `http://ropercenter.cornell.edu/ubs-index-investor-optimism/`. We use the mean point forecast in our estimation and compare these to machine forecasts of the annual CRSP return. Gallup conducted 1,000 interviews of investors during the first two weeks of every month and results were reported on the last Monday of the month. The first survey was conducted on 1998:05. Until 1992:02, the survey was conducted quarterly on 1998:05, 1998:09, and 1998:11. The data on 1998:06, 1998:07, 1998:08, 1998:10, 1998:12, 1999:01, and 2006:01 are missing because the survey was not conducted on these months. We follow Adam, Matveev and Nagel (2021) in starting the sample after 1999:02 due to missing values at the beginning of the sample. For each month when the survey was conducted, respondents are asked about the return they expect on their own portfolio. The survey question is *"What overall rate of return do you expect to get on your portfolio in the next twelve months?"* Before 2003:05, respondents are also asked about the return they expect from an investment in the stock market during the next 12 months. The survey question is *"Thinking about the stock market more generally, what overall rate of return do you think the stock market will provide investors during the coming twelve months?"* For each month, we calculate the average expectations of returns on their own portfolio and returns on the market index. When calculating the average, survey respondents are weighted by the weight factor provided in the survey. We exclude extreme observations where a respondent reported expected returns higher than 95% or lower than -95% on either their own portfolio or the market index.

In order to construct a consistent measure of stock market return expectations over the entire sample period, we impute missing market return expectations using the fitted values from two regressions. First, we impute missing values during 1999:02-2005:12 and 2006:02-2007:10 with the fitted value from regressing expected market returns on own portfolio expectations contemporaneously, where the regression is estimated using the part of the sample where both are available. Second, we impute the one missing observation in both market and own portfolio return expectations for 2006:01 with the fitted value from regressing the market return expectations on the lagged own portfolio return expectations, where the coefficients are estimated using part of the sample where both are available, and the fitted value combines the estimated coefficients with lagged own portfolio expectations data from 2005:12. Following Nagel and Xu (2023), we assume that the forecasted stock market return includes dividends and capture expectations about annual simple net stock returns $\mathbb{F}_t[r_{t+12}]$. To obtain survey expectations of annual log returns $\mathbb{F}_t[\ln(1+r_{t+12})]$ from a survey expectation of annual net simple returns $\mathbb{F}_t[r_{t+12}]$, we use the approximation $\mathbb{F}_t[\ln(1 + r_{t+12})] \approx \ln(1 + \mathbb{F}_t[r_{t+12}])$. After applying all the procedures, the Gallup market return expectations series spans the periods 1999:02 to 2007:10. The data were downloaded on August 1st, 2024.

We take a stand on the information set of respondents when each forecast was made, and we assume that respondents could have used all data released before they completed the survey. Since interviews are in the first two weeks of a month (e.g., February), we conservatively set the response deadline for the machine forecast to be the first day of the survey month (e.g., February 1), implying that we allow the machine to use information only up through the end of the previous month (e.g., through January 31st). This ensures that the machine only sees information that would have been available to all UBS/Gallup respondents for that survey month (February). This approach is conservative in the sense that it handicaps the machine, since all survey respondents who are being interviewed during the next month would have access to more timely information than the machine. Since the survey asks about the "one-year-ahead" we interpret the question to be asking about the forecast period spanning from the current survey month to the same month one year ahead.

CFO Survey The CFO survey is a quarterly survey that asks respondents about their expectations for the S&P 500 return over the next 12 months, obtained from `https://www.richmondfed.org/-/media/RichmondFedOrg/research/national_economy/cfo_survey/current_historical_cfo_data.xlsx`. We use the mean point forecast for the value of the "most likely" future stock return in our estimation. More specifically, the survey asks the respondent "*over the next 12 months, I expect the average annual S&P 500 return will be: Most Likely: I expect the return to be: ___%*". Mean point forecasts before 2020:Q3 are available in column `sp_1_exp` of sheet `through_Q1_2020`; mean point forecasts from 2020:Q3 and onward are available in column `sp_12moexp_2` of sheet `CFO_SP500`. Following Nagel and Xu (2023), we assume that the forecasted S&P 500 return includes dividends and capture expectations about annual simple net stock returns $\mathbb{F}_t[r_{t+12}]$. To obtain survey expectations of annual log returns $\mathbb{F}_t[\ln(1+r_{t+12})]$ from a survey expectation of annual net simple returns $\mathbb{F}_t[r_{t+12}]$, we use the approximation $\mathbb{F}_t[\ln(1+r_{t+12})] \approx \ln(1+\mathbb{F}_t[r_{t+12}])$. The CFO survey panel includes firms that range from small operations to Fortune 500 companies across all major industries. Respondents include chief financial officers, owner-operators, vice presidents and directors of finance, and others with financial decision-making roles. The CFO panel has 1,600 members as of December 2022. As for the SOC, we take a stand on the information set of respondents when each forecast was made, and we assume that respondents could have used all data released before they completed the survey. Because the CFO survey releases quarterly forecasts at the end of each quarter, we conservatively set the response deadline for the machine forecast to be the first day of the last month of each quarter (e.g., March 1). The data spans the periods 2001:Q4 to 2023:Q4. The data were downloaded on March 20th, 2024.

Livingston Survey We obtained the Livingston Survey S&P500 index forecast (SPIF) from the Federal Reserve Bank of Philadelphia, URL: `https://www.philadelphiafed.org/surveys-and-data/real-time-data-research/livingston-historical-data`, and use the mean values in our forecasting models. Our sample spans 1947:06 to 2023:06. The forecast series were downloaded on January 24, 2024.

The survey provides semi-annual forecasts on the level of the S&P 500 index. Participants are asked to provide forecasts for the level of the S&P 500 index for the end of the current survey month, 6 months ahead, and 12 months ahead. We use the mean of the respondents' forecasts each period, where the sample is based on about 50 observations. Most of the survey participants are professional forecasters with "formal and advanced training in economic theory and forecasting and use econometric models to generate their forecasts." Participants receive questionnaires for the survey in May and November, after the Consumer Price Index (CPI) data release for the previous month. All forecasts are typically submitted by the end of the respective month of May and November. The results of the survey are released near the end of the following month, on June and December of each calendar year. The exact release dates are available on the Philadelphia Fed website, at the header of each news release. We take a stand on the information set of the respondents when each forecast was made by assuming that respondents could have used all data released before they completed the survey. Since all forecasts are typically submitted by the end of May and November of each calendar year, we set the response deadline for the machine forecast to be the first day of the last month of June and December, implying that we allow the machine to use information only up through the end of the May and November.

We follow Nagel and Xu (2022) in constructing one-year stock price growth expectations from the level forecasts. Starting from June 1992, we use the ratio between the 12-month level forecast ($\texttt{SPIF\_12M}_t$) and 0-month level nowcasts ($\texttt{SPIF\_ZM}_t$) of the S&P 500 index. Before June 1992, the 0-month nowcast is not available. Therefore we use the annualized ratio between the 12-month ($\texttt{spi12}_t$) and 6-month ($\texttt{spi6}_t$) level forecast of the S&P 500 index

$$\mathbb{F}_t^{(Liv)}\left[\frac{P_{t+12}}{P_t}\right] \approx \begin{cases} \frac{\mathbb{F}_t^{(Liv)}[P_{t+12}]}{\mathbb{F}_t^{(Liv)}[P_t]} = \frac{\texttt{SPIF\_12M}_t}{\texttt{SPIF\_ZM}_t} & \text{if } t \geq 1992M6 \\ \left(\frac{\mathbb{F}_t^{(Liv)}[P_{t+12}]}{\mathbb{F}_t^{(Liv)}[P_{t+6}]}\right)^2 = \left(\frac{\texttt{spi12}_t}{\texttt{spi6}_t}\right)^2 & \text{if } t < 1992M6 \end{cases} \tag{A.1}$$

where P_t is the S&P 500 index and t indexes the survey's response deadline. To obtain a survey expectation of the log change in price growth we use the approximation:

$$\mathbb{F}_t(\Delta p_{t+12}) \approx \ln(\mathbb{F}_t[P_{t+12}]) - \ln(P_t)$$

Finally, price growth expectations from the Livingston survey have been adjusted with the dividend yield to obtain return expectations. We apply the adjustment used in Nagel and Xu (2022) to convert them into return forecasts by adding a time-varying dividend yield:

$$\mathbb{F}_t^{Liv}[r_{t+h}^D] = \mathbb{F}_t^{Liv}\left[\frac{P_{t+1}}{P_t}\right] + \frac{D_t}{P_t}\mathbb{F}_t^{Liv}\left[\frac{D_{t+1}}{D_t}\right] - 1 \tag{A.2}$$

where we set expected dividend growth $\mathbb{F}_t^{Liv}\left[\frac{D_{t+1}}{D_t}\right]$ to 1.064, the sample average of S&P annual dividend growth over the post-war period 1946-2020 (Nagel and Xu (2022)). This adjustment allows us to express all survey inputs to the Kalman filter to be expressed in the same units, i.e., stock returns (with dividends).

Survey of Consumers (SOC) We use the Survey of Consumers (SOC) probabilistic forecast to impute a quantitative point forecast of stock returns using a linear regression of CFO point forecasts for returns onto the SOC probability of a price increase. The SOC asks respondents about the percent chance that an investment will "increase in value in the year ahead." We interpret this as asking about the ex dividend value, i.e., about price price growth. The CFO survey is conducted quarterly, where the survey quarters span 2001:Q4 to 2023:Q4. The SOC survey is conducted monthly, where survey months span 2002:06 to 2023:12. Since the CFO is a quarterly survey, the regression is estimated in real-time over a quarterly overlapping sample. Since the CFO survey is conducted during the last month of the quarter while the SOC is conducted monthly, we align the survey months between CFO and SOC by regressing the quarterly CFO survey point forecast with the qualitative SOC survey response during the last month of the quarter. Since the SOC survey question is interpreted as asking about S&P 500 price growth while the CFO survey question asks about stock returns including dividends, we follow Nagel and Xu (2022) in subtracting the current dividend yield of the CRSP value weighted index from the CFO variable before running the regression. After estimating the regression, we then add back the dividend yield to the fitted value to obtain an imputed SOC point forecast of stock returns including dividends.

Specifically, at time t, we assume that the CFO forecast of stock returns, $\mathbb{F}_t^{\mathrm{CFO}}[r_{t,t+4}]$, minus the current dividend yield, D_t/P_t, is related to the contemporaneous SOC probability of an increase in the stock market next year, $P_{t,t+4}^{\mathrm{SOC}}$, by:

$$\mathbb{F}_t^{\mathrm{CFO}}[r_{t,t+4}] - D_t/P_t = \beta_0 + \beta_1 P_{t,t+4}^{\mathrm{SOC}} + \epsilon_t.$$

The final imputed SOC point forecast is constructed as $\mathbb{F}_t^{\mathrm{SOC}}[r_{t,t+4}] = \hat{\beta}_0 + \hat{\beta}_1 P_{t,t+4}^{\mathrm{SOC}} + D_t/P_t$. We first estimate the coefficients of the above regression over an initial overlapping sample of 2002:Q2 to 2004:Q4, where the quarterly observations from the CFO survey is regressed on the SOC survey responses from the last month of each calendar quarter. Using the estimated coefficients and the SOC probability from 2005:03 gives us the point forecast of the one-year stock return from 2005:Q1 to 2006:Q1. We then re-estimate this equation, recursively, adding one quarterly observation to the end of the sample at a time, and storing the fitted values. This results in a time series of SOC point forecasts $\mathbb{F}_t^{\mathrm{SOC}}[r_{t,t+4}]$ spanning 2005:Q1 to 2023:Q4.

Conference Board (CB) Survey We use the Conference Board (CB) categorical forecast to impute a quantitative point forecast of stock returns by linearly projecting CFO point forecasts for returns onto the CB probability that stock prices will increase. Respondents provide the categorical belief of whether they expect stock prices to "increase," "decrease," or stay the "same" over the next year. We interpret this as a forecast of S&P 500 price growth. Since the survey asks respondents about stock prices in the "year ahead," we interpret the question to be asking about the forecast period from the end of the current survey month to the end of the same month one year ahead. To convert these categorical responses into a quantitative variable, we construct $P_{t,t+4}^{\mathrm{CB}}$ as the ratio of the proportion of respondents who answer "increase" to the proportion of respondents who answer either "decrease" or "same."

The survey uses an address-based mail sample design. Questionnaires are mailed to households on or about the first of each month. Survey responses flow in throughout the collection period, with the sample close-out for preliminary estimates occurring around the 18th of the month. Any responses received after then are used to produce final estimates for the month, which are published with the following month's data. Conversations with those knowledgeable about the survey suggested that most panelists respond early. Any responses received after around the 20th of the month–regardless of when they are filled out–are included in the final (but not preliminary) numbers. We take a stand on the information set of the respondents when each forecast was made by assuming that respondents could have used all data released before they completed the survey. Since questionnaires reach households on or about the first of each month (e.g., February 1st) and most respondents respond early, we conservatively set the response deadline for the machine forecast to be the first day of the survey month (e.g., February 1st), implying that we allow the machine to use information only up through the end of the previous month (e.g., January 31st). We use CB responses from 1987:04–2025:08, with data downloaded on 9/13/2025.

Since the CB survey is about price growth, whereas the CFO survey asks about stock returns including dividends, we follow Nagel and Xu (2022) in subtracting the contemporaneous dividend yield of the CRSP value-weighted index from the CFO variable before running the regression. After estimating the regression, we then add back the dividend yield to the fitted values to obtain an imputed CB point forecast of stock returns including dividends. Specifically, at time t, we assume that the CFO forecast of stock returns $\mathbb{F}_t^{\text{CFO}}[r_{t,t+4}]$, minus the current dividend yield D_t/P_t, is related to the contemporaneous CB ratio $P_{t,t+4}^{\text{CB}}$ by

$$\mathbb{F}_t^{\text{CFO}}[r_{t,t+4}] - \frac{D_t}{P_t} = \beta_0 + \beta_1 P_{t,t+4}^{\text{CB}} + \varepsilon_t.$$

The final imputed CB point forecast of stock returns is then constructed as $\mathbb{F}_t^{\text{CB}}[r_{t,t+4}] = \hat{\beta}_0 + \hat{\beta}_1 P_{t,t+4}^{\text{CB}} + D_t/P_t$.

The CFO survey is conducted quarterly, while the CB survey is conducted monthly. Since CFO data is quarterly, we align the two surveys by using the CB responses from the last month of each quarter. We first estimate the regression over an initial overlapping sample of 2001:Q4–2004:Q4. Using the fitted coefficients and the CB survey response from March 2005 provides the out-of-sample forecast of the one-year stock return for 2005:Q1–2006:Q1. We then re-estimate the regression recursively, adding one quarter at a time and storing the fitted values. This yields a time series of CB point forecasts $\mathbb{F}_t^{\text{CB}}[r_{t,t+4}]$ spanning the forecast testing period of 2005:Q1–2023:Q4.

Survey Deadlines

Filtered Investor Expected Return Series (FI)

State variable is a latent h-month ahead expected stock return capturing investors' subjective beliefs $S_t \equiv \mathbb{F}_t[r_{t+h}^D]$, which evolves according to an AR(1) state equation $S_t = C(\Theta) +$

Table A.2: Quarterly Survey and Machine Forecast Deadlines

Forecast	Survey Deadline	Machine Deadline	Notes
IBES	14-20th of middle month	14th of middle month	Use IBES "Statistical period"
Gallup/UBS	End of week 2 of last month	1st day of last month	Survey conducted during first 2 weeks of each month
CFO	End of week 2 of last month	1st day of last month	Survey conducted during 2 weeks starting early in the last month
Livingston	First day of Jun & Dec	First day of Jun & Dec	Survey conducted during last week of May & Nov
SOC	Last day of last month	1st day of last month	Survey conducted throughout last month
CB	18th of last month	1st day of last month	Preliminary estimates close on the 18th

Notes: Survey deadlines correspond to when respondents typically submit forecasts. Machine deadlines are set conservatively to ensure the algorithm only uses information available at or before the survey deadline.

$T(\Theta)S_{t-1} + R(\Theta)\varepsilon_t$

$$S_t = \underbrace{\alpha}_{C(\Theta)} + \underbrace{\rho}_{T(\Theta)} S_{t-1} + \underbrace{\sigma_\varepsilon}_{R(\Theta)} \varepsilon_t \tag{A.3}$$

where C, T, R are matrices of the model's primitive parameters $\Theta = (\alpha, \rho, \sigma_\varepsilon)'$. ε_t is an innovation to the latent expectation that was unpredictable from the point of view of the forecaster. α is the intercept, ρ is the persistence, and σ_ε is the standard deviation of the latent innovation error. Observation equation takes the form $X_t = D + ZS_t + Uv_t$

$$\underbrace{\begin{bmatrix} \mathbb{F}_t^{Gallup/UBS}[r_{t+h}^D] \\ \mathbb{F}_t^{CFO}[r_{t+h}^D] \\ \mathbb{F}_t^{Liv}[r_{t+h}^D] \\ \mathbb{F}_t^{SOC}[r_{t+h}^D] \end{bmatrix}}_{X_t} = D_t + Z_t S_t + U_t v_t \tag{A.4}$$

where $h = 12$ months is a fixed forecast horizon. The observation vector X_t contains measures of survey expected returns over the next h periods. v_t is a vector of observation errors with standard deviations in the diagonal matrix U. Z and D are parameters that have been set to 1s and 0s, respectively. We use the Kalman filter to estimate the remaining parameters $\alpha, \rho, \sigma_\varepsilon, U$. We note that the coefficient matrix Z_t and vectors D_t and U_t depend on t because some of the survey series are not available for the full sample or are available at different sampling intervals. As a result, the state space estimation uses different measurement equations to include observations for these series when they are available and exclude them when they are missing.

All survey observations are interpreted as forecasts of returns (with dividends). When survey questions refer only to expected price growth, we apply the adjustment used in Nagel and Xu (2022) to convert them into return forecasts by adding a time-varying dividend yield:

$$\mathbb{F}_t^{Liv}[r_{t+h}^D] = \mathbb{F}_t^{Liv}\left[\frac{P_{t+1}}{P_t}\right] + \frac{D_t}{P_t}\mathbb{F}_t^{Liv}\left[\frac{D_{t+1}}{D_t}\right] - 1 \tag{A.5}$$

where we set expected dividend growth $\mathbb{F}_t^{Liv}\left[\frac{D_{t+1}}{D_t}\right]$ to 1.064, the sample average of S&P annual dividend growth over the post-war period 1946-2020 (Nagel and Xu (2022)). For the Livingston survey, which elicits expected price growth, we add a smoothed dividend yield to produce a return forecast. For SOC, after converting qualitative responses into point forecasts, we apply the same dividend yield adjustment. Gallup/UBS and CFO surveys are interpreted directly as return forecasts and do not require further adjustment. See Section 3.4 for more details.

Perceived Risk

We follow Nagel and Xu (2022) in constructing alternative measures of perceived risk in the stock market.

Perceived Variance The CFO survey is a quarterly survey that asks respondents about their expectations for the S&P 500 return over the next 12 months, obtained from `https://www.richmondfed.org/research/national_economy/cfo_survey`. In the CFO survey, respondents are asked to provide their expectation of the 10th and 90th percentile of the stock market return distribution over the next 12 months. The survey question for the 10th percentile is "*Over the next 12 months, I expect the average annual S&P 500 return will be: Worst Case: There is a 1-in-10 chance the actual return will be less than: ___*". The survey question for the 90th percentile is "*Over the next 12 months, I expect the average annual S&P 500 return will be: Best Case: There is a 1-in-10 chance the actual return will be greater than: ___*". We take the difference between the mean expectations of 90th and 10th percentiles of returns. To convert this difference into a measure of subjective variance, we take the square of the difference and divide by the square of 2.56. The transformation would be accurate if the stock market return distribution was normal. The data were downloaded on March 20th, 2024.

Crash Confidence Index The Crash Confidence Index measures the percent of respondents that believe there is less than 10% chance of a stock market crash in the next six months. Survey respondents are asked the question: "What do you think is the probability of a catastrophic stock market crash in the U.S., like that of October 28, 1929 or October 19, 1987, in the next six months, including the case that a crash occurred in the other countries and spreads to the U.S.? (An answer of 0% means that it cannot happen, an answer of 100% means it is sure to happen.)" The survey provides two indices based on two separate samples: a sample of wealthy individual investors, and another sample of institutional investors. The sample is drawn randomly from a list of individual and institutional investors purchased from InfoUSA (Data Axle). We use the negative of each series to proxy for individual and institutional investors' perceived risk. On average, the sample size is about one hundred respondents for each six-month interval. The sample spans 1989:10-2021:12 for Crash Confidence Index (Institutions), with semi-annual observations over 1989:10-2001:04 on Q2 and Q4 of each calendar year; 1989:10, 1996:10, and 1999:04-2021:12 for Crash Confidence Index (Individuals), with semi-annual observations over 1999:04-2001:04 on Q2 and Q4 of each calendar year. Starting 2001:07, the survey series report monthly six-month trailing averages of the monthly survey values. For example, the reported number for 2018:01 is an average of survey results from 2017:08 to 2018:01. We follow Nagel and Xu (2023) in treating the 6 month moving average of the 6 month ahead crash confidence measure as a proxy for 12-month ahead uncertainty. The data were downloaded from the Yale International Center for Finance website on March 20, 2024 at URL: `https://som.yale.edu/centers/international-center-for-finance/data/stock-market-confidence-indices`.

Financial Uncertainty Index The Financial Uncertainty Index is constructed following Jurado, Ludvigson and Ng (2015) and reflects a measure of objective uncertainty about real activity over a 12-month horizon. The data has been downloaded on September 25, 2024 from the author's website: `https://www.sydneyludvigson.com/s/MacroFinanceUncertainty_202408Update.zip`.

Factor Data

This section describes the data sources for the factor returns used in Table 5. The sample period for all factors is 2005:Q1 to 2023:Q4.

CAPM market factor The CAPM market factor is the excess return on the CRSP value-weighted market portfolio, computed as the return on all NYSE, AMEX, and NASDAQ stocks minus the one-month Treasury bill rate. The CRSP data were obtained through WRDS and downloaded on August 13, 2025.

Size and value factors (SMB and HML) The size factor (SMB, "Small Minus Big") is the average return on three small-cap portfolios minus the average return on three large-cap portfolios. The value factor (HML, "High Minus Low") is the average return on two high book-to-market portfolios minus the average return on two low book-to-market portfolios. Both series were downloaded on August 13, 2025 from the Fama/French 3 Factors file at `https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`.

Profitability and investment factors (RMW and CMA) The profitability factor (RMW, "Robust Minus Weak") is the average return on two robust operating profitability portfolios minus the average return on two weak operating profitability portfolios. The investment factor (CMA, "Conservative Minus Aggressive") is the average return on two low investment portfolios minus the average return on two high investment portfolios. Both series were downloaded on August 13, 2025 from the Fama/French 5 Factors file at `https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`.

Natural Language Processing: LDA Data

These data are used as inputs into the machine learning forecasts. The database for our Latent Dirichlet Allocation (LDA) analysis contains around one million articles published in *Wall Street Journal* between January 1984 to June 2022. The current vintage of the results reported here is based a randomly selected sub-sample of 200,000 articles over the same period, one-fifth size of the entire database. The sample selection procedure follows Bybee et al. (2021). First, we remove all articles prior to January 1984 and after June 2024 and exclude articles published in weekends. Second, we exclude articles with subject tags associated with obviously non-economic content such as sports. Third, we exclude articles with the certain headline patterns, such as those associated with data tables or those corresponding to regular sports, leisure, or books columns. We filter the articles using the same list of exclusions provided by Bybee et al. (2021). Last, we exclude articles with less than 100 words.

Processing of texts The processing of the texts can be summarized in the following five steps.

1. Tokenization: parse each article's text into a white-space-separated word list retaining the article's word ordering.

2. We drop all non-alphabetical characters and set the remaining characters to lower-case, remove words with less than 3 letters, and remove common stop words and URL-based terms. We use a standard list of stop words from the Python library *gensim.parsing.preprocessing*.

3. Lemmatization and Stemming: lemmatization returns the original form of a word using external dictionary *Textblob.Word* in Python and based on the context of the word. For instance, as a verb, "went" is converted to"go". Stemming usually refers to a heuristic process that remove the trailing letters at the end of the words, such as from "assesses" to "assess', and "really" to "real". We use the Python library *Textblob.Word* to implement the lemmatization and *SnowballStemmer* for the stemming. The results are not very sensitive to the particular Python packages being used.

4. From the first three steps, we obtain a list of uni-grams which are a list of singular words. For example, "united" and "states" are uni-grams from "united states". From the list of uni-grams, we generate a set of bi-grams as all pairs of (ordered) adjacent uni-grams. For example, "united states" together is one bi-gram. We then exclude uni-grams and bi-grams appearing in less than 0.1% of articles.

5. Last, we convert an article's word list into a vector of counts for each uni-gram and bi-gram. For example, the vector of counts $[5, 7, 2]$ corresponds to the number of times the words $[''federal'', ''reserve'', ''bank'']$ appear in the article.

Data Inputs for Machine Learning Algorithm

Macro Data Surprises

These data are used as inputs into the machine learning forecasts. We obtain median forecasts for GDP growth (Q/Q percentage change), core CPI (Month/Month change), unemployment rate (percentage point), and nonfarm payroll (month/month change) from the Money Market Service Survey. The median market survey forecasts are compiled and published by the Money Market Services (MMS) the Friday before each release. We apply the approach used in Bauer and Swanson (2023) and define macroeconomic data surprise as the actual value of the data release minus the median expectation from MMS on the Friday immediately prior to that data release. The GDP growth forecasts are available quarterly from 1990:Q1 to 2023:Q4. The core CPI forecast is available monthly from July 1989 to Dec 2023. The median forecasts for the unemployment rate and nonfarm payrolls are available monthly from Jan 1980 to Dec 2023, and Jan. 1985 to Dec 2023, respectively. All survey forecasts were downloaded from Haver Analytics and the Bloomberg Terminal on July 20, 2025. To pin down the timing of when the news was actually released we follow the published tables of releases from the Bureau of Labor Statistics (BLS), discussed below.

FOMC Surprises

FOMC surprises are defined as the changes in the current-month, 1, 2, 6, 12, and 24 month-ahead federal funds futures (FFF) contract rate and changes in the 1, 2, 4, and 8 quarter-ahead Eurodollar (ED) futures contract rate, from 10 minutes before to 20 minutes after each U.S. Federal Reserve Federal Open Market Committee (FOMC) announcement. The data on FFF and ED were downloaded on July 9, 2025. When benchmarking against a survey, we use the last FOMC meeting before the survey deadline to compute surprises. For surveys that do not have a clear deadline, we compute surprises using from the last FOMC in the first month of the quarter. When benchmarking against moving average, we use the last FOMC meeting before the end of the first month in each quarter to compute surprises.

Real-Time Macro Data

This section gives details on the real time macro data inputs used in the machine learning forecasts. A subset of these series are used in the structural estimation. 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.[2] 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.

Table A.3 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

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

vintage because most variables have vintages through the present.[3] Table A.3 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.3: 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

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

Table A.2 (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

Table A.2 (Cont'd)

No.	Short Name	Source	Tran	Description	First Vintage
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

Table A.2 (Cont'd)

No.	Short Name	Source	Tran	Description	First Vintage
			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
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

Table A.2 (Cont'd)

No.	Short Name	Source	Tran	Description	First Vintage
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

Daily Financial Data

Daily Data and construction of daily factors These data are used in the machine learning forecasts. 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 31-Dec-2023 (16,126 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 and standardized before performing factor estimation (see generic description below). We re-estimate factors at each date in the sample recursively over time using the entire history of data available in real time prior to each out-of-sample forecast.

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

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

Table A.4: List of Daily Financial Dataset Variables

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

Table A.3 (Cont'd)

No.	Short Name	Source	Tran	Description
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

Table A.3 (Cont'd)

No.	Short Name	Source	Tran	Description
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.3 (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 (%)

Table A.3 (Cont'd)

No.	Short Name	Source	Tran	Description
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

The LDA Model The LDA model Blei, Ng and Jordan (2003) essentially achieves substantial dimension reduction of the word distribution of each article using the following assumptions. We assume a factor structure on the vectors of word counts. Each factor is a topic and each article is a parametric distribution of topics, specified as follows,

$$\underbrace{\overbrace{w_i}^{V\times 1}}_{\text{word dist of article } i} \sim \text{Mult}\left(\underbrace{\overbrace{\Phi'}^{V\times K}}_{\text{topic-word dist.}}, \underbrace{\overbrace{\theta_i}^{K\times 1}}_{\text{topic dist.}}, \underbrace{N_i}_{\#\text{ of words}}\right) \tag{A.6}$$

where Mult is the multinomial distribution. In the above equation, w_i is a vector of word counts of each unique term (uni-gram or bi-gram) in article i, whose size is equal to the number of unique terms V. K is the number of factors in article i. In the estimation, we assume $K = 180$ following Bybee et al. (2021). Φ is a matrix sized $K \times V$, whose kth row and vth column is equal to the probability of the unique term v showing up in topic k. θ_i stores the weights of all k topics contained in article i, which sum up to one. Dimension reduction is achieved as long as $K << V$ (the number of topics are significantly smaller than the number of unique terms). More specifically, it reduces the dimension from $T \times V$ to $T \times K$ (the size of θ) + $K \times V$ (the size of Φ).

Real-time news factors. We also generate real-time news factors for each month t starting from January 1991. In theory, we could train the LDA model using each real-time monthly vintage but it is computationally challenging. Instead, we simplify the procedure by training the LDA model using quarterly vintages t, $t+3$, $t+6$, etc, and use the LDA model parameters estimated at t to filter news paper articles within the quarter and generate news factors for those months. More specifically, given every article's word distribution $w_{i,t+s}$,for $s = 0, 1, 2$, and the estimated real-time topic-word distribution parameters $\hat{\Phi}_t$ using articles till date t, one can obtain the filtered topic distribution of each article $\hat{\theta}_{i,t+s}$, as follows,

$$\underbrace{\overbrace{w_{i,t+s}}^{V\times 1}}_{\text{word dist of article } i \text{ at time } t+s} \sim \text{Mult}\left(\underbrace{\overbrace{\hat{\Phi}'}^{V\times K}}_{\text{topic-word dist.}}, \underbrace{\overbrace{\hat{\theta}_{i,t+s}}^{K\times 1}}_{\text{topic dist.}}, \underbrace{N_{i,t+s}}_{\#\text{ of words}} \right). \tag{A.7}$$

LDA Estimation We use the built-in LDA model estimation toolbox in the Python library `https://pypi.org/project/gensim/`Gensim to implement the model estimation. The model requires following initial inputs and parameters and it is estimated using Bayesian methods.[4]

1. We create a document-term matrix $\mathbf{W}$ as a collection of w_i for all articles i in the sample. The number of rows in $\mathbf{W}$ is equal to the number of articles in our sample and the number of columns in $\mathbf{W}$ is equal to the number of unique uni-gram and bi-grams (after being filtered) across all articles. The matrix $\mathbf{W}$ is used as an input for the LDA model estimation. We then follow Bybee et al. (2021) and set the number of topics K to be 180.[5]

2. In the Python library Gensim, the key parameters of the LDA estim are α and β.With a higher value of α, the documents are composed of more topics. With a higher values of β, each topic contains more terms (uni- or bi-grams). In the implementations, we do not impose any explicit restrictions on initial values of those parameters and set

[4]In theory, maximum-likelihood estimation is possible but it is computationally challenging.
[5]The authors used Bayesian criteria to find 180 to be an optimal number of topics.

them to be “auto”. These two parameters, alongside Φ' and $\{\theta_i\}_i$, are estimated by the toolbox from Python library `https://pypi.org/project/gensim/`Gensim.

Real-time LDA Factors With the estimated topic weights $\theta_{i,t}$ of each article i from the LDA model, we fruther construct time series of the overall news attention to each topic, or a news factor. The value of the topic k at time t is the average weights of topic k of all articles published at t, specified as follows,

$$F_{k,t} = \frac{\sum_i \hat{\theta}_{i,k,t}}{\#\text{ of articles at } t} \tag{A.8}$$

for all topics k.

Daily LDA Factors We construct daily LDA factors by aggregating all articles published on each calendar day. The value of topic k at day t is the average weights of topic k across all articles published that day.

Machine Learning

Machine Algorithm Details

The basic dynamic algorithm follows the six step approach of Bianchi et al. (2022) of 1. Sample partitioning, 2. In-sample estimation, 3. Training and cross-validation, 4. Grid reoptimization, 5. Out-of-sample prediction, and 6. Roll forward and repeat. We refer the interested reader to that paper for details and discuss details of the implementation here only insofar as they differ.

At time t, a prior sample of size $\dot{T}$ is partitioned into two sub-sample windows: a *training sample* consisting of the first T_E observations, and a hold-out *validation sample* of T_V subsequent observations so that $\dot{T} = T_E + T_V$. The training sample is used to estimate the model subject to a specific set of tuning parameter values, and the validation sample is used for tuning the hyperparameters. The model to be estimated over the training sample is

$$y_{j,t+h} = G^e\left(\mathcal{X}^t, \boldsymbol{\beta}_{j,h,t}\right) + \epsilon_{j,t+h}.$$

where $y_{j,t+h}$ is a time series indexed by j whose value in period $h \geq 1$ the machine is asked to predict at time t, $\mathcal{X}^t$ is the history of a large input dataset of right-hand-side variables (including the intercept) from time $1, ..., t$, and $G^e(\cdot)$ is a machine learning estimator that can be represented by a (potentially) high dimensional set of finite-valued parameters $\boldsymbol{\beta}^e_{j,h,t}$. We consider two estimators for $G^e(\cdot)$: Elastic Net $G^{\text{EN}}(\mathcal{X}^t, \boldsymbol{\beta}^{\text{EN}}_{j,h})$, and Long Short-Term Memory (LSTM) network $G^{\text{LSTM}}(\mathcal{X}^t, \boldsymbol{\beta}^{\text{LSTM}}_{j,h})$. The $e \in \{\text{EN}, \text{LSTM}\}$ superscripts on $\boldsymbol{\beta}$ indicate that the parameters depend on the estimator being used (See the next section for a description of EN and LSTM). $\mathcal{X}_t$ always denotes the most recent data that would have been in real time prior to the date on which the forecast was submitted. To ensure that the effect of each variable in the input vector is regularized fairly during the estimation, we standardize the elements of $\mathcal{X}_t$ such that sample means are zero and sample standard deviations are unity. 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.

The parameters $\boldsymbol{\beta}^e_{j,h,t}$ are estimated by minimizing the mean-square loss function over the training sample with L_1 and L_2 penalties

$$L(\boldsymbol{\beta}^e_{j,h,t}, \mathcal{X}^{T_E}, \boldsymbol{\lambda}^e_t) \equiv \underbrace{\frac{1}{T_E}\sum_{\tau=1}^{T_E}\left(y_{j,\tau+h} - G^e\left(\mathcal{X}^\tau, \boldsymbol{\beta}^e_{j,h,t}\right)\right)^2}_{\text{Mean Square Error}} + \underbrace{\lambda^e_{1,t}\sum_{k=1}^{K}\left|\boldsymbol{\beta}^e_{j,h,t,k}\right|}_{L_1 \text{ Penalty}} + \underbrace{\lambda^e_{2,t}\sum_{k=1}^{K}(\boldsymbol{\beta}^e_{j,h,t,k})^2}_{L_2 \text{ Penalty}}$$

where $\mathcal{X}^{T_E} = (\mathcal{X}'_1, \dots, \mathcal{X}'_{T_E})'$ is the vector containing all observations in the training sample of size T_E. The estimated $\boldsymbol{\beta}^e_{j,h,t}$ is a function of the data $\mathcal{X}^{T_E}$ and a non-negative regularization parameter vector $\boldsymbol{\lambda}^e_t = \left(\lambda^e_{1,t}, \lambda^e_{2,t}, \boldsymbol{\lambda}^{LSTM}_{0,t}\right)'$ where $\boldsymbol{\lambda}^{LSTM}_{0,t}$ is a set of hyperparameters only relevant when using the LSTM estimator for $G^e(\cdot)$ (see below). For the EN case there are only two hyperparameters, which determine the optimal shrinkage and sparsity of the time t machine specification. The regularization parameters $\boldsymbol{\lambda}^e_t$ are estimated by minimizing the mean-square loss over pseudo-out-of-sample forecast errors generated from rolling regressions through the validation sample:

$$\widehat{\boldsymbol{\lambda}}^{EN}_t, \widehat{T}_E, \widehat{T}_V = \underset{\boldsymbol{\lambda}^{EN}_t, T_E, T_V}{argmin}\left\{\frac{1}{T_V - h}\sum_{\tau=T_E}^{T_E+T_V-h}\left(y_{j,\tau+h} - G^{EN}(\mathcal{X}_\tau, \widehat{\boldsymbol{\beta}}^{EN}_{j,h,\tau}(\mathcal{X}^\tau, \boldsymbol{\lambda}^{EN}_t))\right)^2\right\}$$

$$\widehat{\boldsymbol{\lambda}}^{LSTM}_t, \widehat{T}_E, \widehat{T}_V = \underset{\boldsymbol{\lambda}^{LSTM}_t, T_E, T_V}{argmin}\left\{\frac{1}{T_V - h}\sum_{\tau=T_E}^{T_E+T_V-h}\left(y_{j,\tau+h} - G^{LSTM}(\mathcal{X}_\tau, \widehat{\boldsymbol{\beta}}^{LSTM}_{j,h,\tau}(\mathcal{X}^\tau, \boldsymbol{\lambda}^{LSTM}_t))\right)^2\right\}$$

where $\widehat{\boldsymbol{\beta}}^e_{j,h,\tau}(\cdot)$ for $e \in \{\text{EN}, \text{LSTM}\}$ is the time τ estimate of $\boldsymbol{\beta}^e_{j,h}$ given $\boldsymbol{\lambda}^e_t$ and data through time τ in a training sample of size T_E. Denote the combined final estimator $\widehat{\boldsymbol{\beta}}^e_{j,h,t}(\mathcal{X}^t, \widehat{\boldsymbol{\lambda}}^e_t)$, where the regularization parameter $\widehat{\boldsymbol{\lambda}}^e_t$ is estimated using cross-validation dynamically over time. Note that the algorithm also asks the machine to dynamically choose both the optimal training window $\widehat{T}_E$ and the optimal validation window $\widehat{T}_V$ by minimizing the pseudo-out-of-sample MSE.

The estimation of $\widehat{\boldsymbol{\beta}}^e_{j,h,t}(\mathcal{X}^t, \widehat{\boldsymbol{\lambda}}^e_t)$ is repeated sequentially in rolling subsamples, with parameters estimated from information known at time t. Note that the time t subscripts of $\widehat{\boldsymbol{\beta}}^e_{j,h,t}$ and $\widehat{\boldsymbol{\lambda}}^e_t$ 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 denote the time t machine belief about $y_{j,t+h}$ as $\mathbb{E}^e_t[y_{j,t+h}]$, defined by

$$\mathbb{E}^e_t[y_{j,t+h}] \equiv G^e\left(\mathcal{X}^t, \widehat{\boldsymbol{\beta}}^e_{j,h,t}(\mathcal{X}^t, \widehat{\boldsymbol{\lambda}}^e_t)\right)$$

Finally, the machine MSE is computed by averaging across the sequence of squared forecast errors in the true out-of-sample forecasts for periods $t = (\dot{T} + h), \ldots, T$ where T is the last period of our sample. The true out-of-sample forecasts used for neither estimation nor tuning is the *testing sub-sample* used to evaluate the model's predictive performance.

On rare occasions, one or more of the explanatory variables used in the machine forecast specification assumes a value that is order of magnitudes different from its historical value. This is usually indicative of a measurement problem in the raw data. We therefore program the machine to detect in real-time whether its forecast is an extreme outlier, and in that case to discard the forecast replacing it with the historical mean. Specifically, at each t, the machine forecast $\mathbb{E}^e_t\left[y_{j,t+h}\right]$ is set to be the historical mean calculated up to time t whenever the former is five or more standard deviations above its own rolling mean over the most recent 20 quarters.

We include the contemporaneous investor forecasts $\mathbb{F}_t\left[y_{j,t+h}\right]$ for the median respondent only, following Bianchi et al. (2022). This procedure allows the machine to capture intangible information due to judgement or private signals. Specifically, we 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 investor forecast of respondent-type X, where this type is the median here:

$$y_{j,t+h} = G^e_{jh}\left(\mathcal{X}^t\right) + \gamma_{jh\mathbb{M}}\mathbb{F}_t\left[y_{j,t+h}\right] + \epsilon_{j,t+h}, \qquad h \geq 1 \tag{A.9}$$

where $\gamma_{jh\mathbb{M}}$ is a parameter to be estimated, and where $G_{jh\mathbb{M}}\left(\mathcal{X}^t\right)$ represents a ML estimator as function of big data. Note that the intercept α_{jh} from Bianchi et al. (2022) gets absorbed into the $G^e_{jh}\left(\mathcal{X}^t\right)$ in LSTM via the outermost bias term.

Elastic Net (EN)

Output from the Elastic Net (EN) estimator is sometimes used as an input into the main Long-Short-Term-Memory model described below. EN combines Least Absolute Shrinkage and Selection Operator (LASSO) and ridge type penalties. The model can be written as:

$$y_{j,t+h} = \mathcal{X}'_{tj}\boldsymbol{\beta}^{\text{EN}}_{j,h} + \epsilon_{j,t+h}$$

where $\mathcal{X}_t = (1, \mathcal{X}_{1t,...,}\mathcal{X}_{Kt})'$ include the independent variable observations $\left(\mathbb{F}_t\left[y_{j,t+h}\right], \mathcal{Z}_{j,t}\right)$ into a vector with "1" and $\boldsymbol{\beta}^{\text{EN}}_{j,h} = \left(\alpha_{j,h}, \beta_{j,h\mathbb{F}}, vec\left(\mathbf{B}_{j,h\mathcal{Z}}\right)\right)' \equiv \left(\beta_0, \beta_1, ...\beta_K\right)'$ collects all the coefficients.

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:

$$\widehat{\boldsymbol{\beta}}^{\text{EN}}_{j,h} = \underset{\beta_0,\beta_1,...,\beta_K}{argmin}\left\{\frac{1}{T_E}\sum_{\tau=1}^{T_E}\left(y_{j,\tau+h} - \mathcal{X}'_{\tau}\boldsymbol{\beta}_{j,h}\right)^2 + \underbrace{\lambda_1\sum_{k=1}^{K}\left|\boldsymbol{\beta}_{j,h,k}\right|}_{\text{LASSO}} + \underbrace{\lambda_2\sum_{k=1}^{K}(\boldsymbol{\beta}_{j,h,k})^2}_{\text{ridge}}\right\}$$

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

In the implementation, the EN estimator is sometimes used as an input into the algorithm using the LSTM estimator. Specifically, we ensure that the machine forecast can only differ from the relevant benchmark if it demonstrably improves the pseudo out-of-sample prediction in the training samples *prior* to making a true out-of-sample forecast. Otherwise, the machine is replaced by the benchmark calculated up to time t. In some cases the benchmark is a investor forecast, in others it could be a historical mean value for the variable. However, for the implementation using LSTM, we also use the EN forecast as a benchmark.

Long Short-Term Memory (LSTM) Network

An LSTM network is a type of Recurrent Neural Network (RNN), which are neural networks used to learn about sequential data such as time series or natural language. In particular, LSTM networks can learn long-term dependencies between across time periods by introducing hidden layers and memory cells to control the flow of information over longer time periods. The general case of the LSTM network with up to N hidden layers is defined as follows. The 0-th layer of the network is the input data: $h_t^0 \equiv \mathcal{X}_t$. This initial value feeds into higher-order hidden layers according to:

$$\underbrace{G^{\text{LSTM}}(\mathcal{X}^t, \boldsymbol{\beta}_{j,h}^{\text{LSTM}})}_{1\times 1} = \underbrace{W^{(yh^N)}}_{1\times D_{h^N}} \underbrace{h_t^N}_{D_{h^N}\times 1} + \underbrace{b_y}_{1\times 1} \quad \text{(Output layer)}$$

$$\underbrace{h_t^n}_{D_{h^n}\times 1} = \underbrace{o_t^n}_{D_{h^n}\times 1} \odot \tanh(\underbrace{c_t^n}_{D_{h^n}\times 1}) \quad \text{(Hidden layer)}$$

$$\underbrace{c_t^n}_{D_{h^n}\times 1} = \underbrace{f_t^n}_{D_{h^n}\times 1} \odot \underbrace{c_{t-1}^n}_{D_{h^n}\times 1} + \underbrace{i_t^n}_{D_{h^n}\times 1} \odot \underbrace{\widetilde{c}_t^n}_{D_{h^n}\times 1} \quad \text{(Final memory)}$$

$$\underbrace{\widetilde{c}_t^n}_{D_{h^n}\times 1} = \tanh(\underbrace{W^{(c^n h^{n-1})}}_{D_{h^n}\times D_{h^{n-1}}} \underbrace{h_t^{n-1}}_{D_{h^{n-1}}\times 1} + \underbrace{W^{(c^n h^n)}}_{D_{h^n}\times D_{h^n}} \underbrace{h_{t-1}^n}_{D_{h^n}\times 1} + \underbrace{b_{c^n}}_{D_{h^n}\times 1}) \quad \text{(New memory)}$$

$$\underbrace{f_t^n}_{D_{h^n}\times 1} = \sigma(\underbrace{W^{(f^n h^{n-1})}}_{D_{h^n}\times D_{h^{n-1}}} \underbrace{h_t^{n-1}}_{D_{h^{n-1}}\times 1} + \underbrace{W^{(f^n h^n)}}_{D_{h^n}\times D_{h^n}} \underbrace{h_{t-1}^n}_{D_{h^n}\times 1} + \underbrace{b_{f^n}}_{D_{h^n}\times 1}) \quad \text{(Forget gate)}$$

$$\underbrace{i_t^n}_{D_{h^n}\times 1} = \sigma(\underbrace{W^{(i^n h^{n-1})}}_{D_{h^n}\times D_{h^{n-1}}} \underbrace{h_t^{n-1}}_{D_{h^{n-1}}\times 1} + \underbrace{W^{(i^n h^n)}}_{D_{h^n}\times D_{h^n}} \underbrace{h_{t-1}^n}_{D_{h^n}\times 1} + \underbrace{b_{i^n}}_{D_{h^n}\times 1}) \quad \text{(Input gate)}$$

$$\underbrace{o_t^n}_{D_{h^n}\times 1} = \sigma(\underbrace{W^{(o^n h^{n-1})}}_{D_{h^n}\times D_{h^{n-1}}} \underbrace{h_t^{n-1}}_{D_{h^{n-1}}\times 1} + \underbrace{W^{(o^n h^n)}}_{D_{h^n}\times D_{h^n}} \underbrace{h_{t-1}^n}_{D_{h^n}\times 1} + \underbrace{b_{o^n}}_{D_{h^n}\times 1}) \quad \text{(Output gate)}$$

where $n = 1, \ldots, N$ indexes each hidden layer. $h_t^n \in \mathbb{R}^{D_{h^n}}$ is the n-th *hidden layer*, where D_{h^n} is the number of *neurons* or *nodes* in the hidden layer. The memory cell c_t^n allows the LSTM network to retain information over longer time periods. The output gate o_t^n controls the extent to which the memory cell c_t^n maps to the hidden layer h_t^n. The forget gate f_t^n controls the flow of information carried over from the final memory in the previous timestep c_{t-1}^n. The input gate i_t^n controls the flow of information from the new memory cell $\widetilde{c}_t^n$. The initial states for the hidden layers $(h_0^n)_{n=1}^N$ and memory cells $(c_0^n)_{n=1}^N$ are set to zeros. $\sigma(\cdot)$ and $\tanh(\cdot)$ are *activation functions* that introduce non-linearities in the LSTM network, applied elementwise. $\sigma : \mathbb{R} \to \mathbb{R}$ is the sigmoid function: $\sigma(x) = (1 + e^{-x})^{-1}$. $\tanh : \mathbb{R} \to \mathbb{R}$ is the hyperbolic tangent function: $\tanh(x) = \frac{e^{2x}-1}{e^{2x}+1}$. The $\odot$ operator refers to elementwise multiplication. $\boldsymbol{\beta}_{j,h}^{\text{LSTM}} \equiv (((vec(W^{(g^n h^{n-1})})', vec(W^{(g^n h^n)})', b'_{g^n})_{g\in\{c,f,i,o\}})_{n=1}^N, vec(W^{(yh^N)})', b_y)'$ are parameters to be estimated. We will refer to parameters indexed with W as *weights*; parameters indexed with b are *biases*.

At each point in time, we estimate the parameters $\boldsymbol{\beta}_{j,h}$ for the LSTM network using Stochastic Gradient Decent (SGD), which is an iterative algorithm for minimizing the loss function and proceeds as follows:

1. *Initialization.* Fix a random seed R and draw a starting value of the parameters $\boldsymbol{\beta}_{j,h}^{(0)}$ randomly, where the superscript (0) in parentheses indexes the iteration for an estimate of $\boldsymbol{\beta}_{j,h}^{\text{LSTM}}$.

   (a) Initialize the input weights $W^{(g^n h^{n-1})} \in \mathbb{R}^{D_{h^n} \times D_{h^{n-1}}}$ for $g \in \{c, f, i, o\}$ using the *Glorot* initializer. Draw randomly from a uniform distribution with zero mean and a variance that depends on the dimensions of the matrix:

   $$W_{ij}^{(g^n h^{n-1})} \overset{iid}{\sim} U\left[-\sqrt{\frac{6}{D_{h^n} + D_{h^{n-1}}}}, \sqrt{\frac{6}{D_{h^n} + D_{h^{n-1}}}}\right]$$

   for each $i = 1, \ldots, D_{h^n}$ and $j = 1, \ldots, D_{h^{n-1}}$.

   (b) Initialize the recurrent weights $W^{(g^n h^n)} \in \mathbb{R}^{D_{h^n} \times D_{h^n}}$ for $g \in \{c, f, i, o\}$ using the *Orthogonal* initializer. Use the orthogonal matrix obtained from the QR decomposition of a $D_{h^n} \times D_{h^n}$ matrix of random numbers drawn from a standard normal distribution.

   (c) Initialize biases $(b_{g^n})_{g\in\{c,f,i,o\}}$, hidden layers h_0^n, and memory cells c_0^n with zeros.

2. *Mini-batches.* Prepare the input data by dividing the training sample into a collection of *mini-batches.*

   (a) Suppose that we have a multi-variate time-series training sample with dimensions (T_E, K) whose time steps t are indexed by $t = 1, \ldots, T_E$ and K is the number of predictors. We transform this training sample into a 3-D tensor with dimensions (N_S, M, K) where

   - N_S = Total number of sequences in training sample
   - M = Sequence length, i.e., number of time steps in each sequence
   - K = Input size, i.e., number of predictors in each time step

   This can be done by creating overlapping sequences from the time series:

   - Sequence 1 contains time steps $1, \ldots, M$

- Sequence 2 contains time steps $2, \dots, M+1$
- Sequence 3 contains time steps $3, \dots, M+2$
- ...
- Sequence $T_E - M$ contains time steps $T_E - M, \dots, T_E - 1$
- Sequence $N_S = T_E - M + 1$ contains time steps $T_E - M + 1, \dots, T_E$

(b) Randomly shuffle the N_S sequences by randomly sampling a permutation of the sequences without replacement.

(c) Partition the N_S shuffled sequences into $\lceil N_S/N_B \rceil$ mini-batches. We partition the N_S sequences in the training sample ((N_S, M, K) tensor) into a list of $\lceil N_S/N_B \rceil$ mini-batches. A mini-batch is a (N_B, M, K)-dimensional tensor containing N_B out of N_S randomly shuffled sequences.[6] Let $B^{(1)}, \dots, B^{\lceil N_S/N_B \rceil}$ denote the list of mini-batches.

- N_S = Total number of sequences in training sample
- N_B = Mini-batch size, i.e., number of sequences in each partition.
- M = Sequence length, i.e., number of time steps in each sequence
- K = Input size, i.e., number of predictors in each time step

3. Repeat until the stopping condition is satisfied ($k = 1, 2, 3, \dots$):

(a) *Dropout.* Apply dropout to the mini-batch. To obtain the n-th hidden layer under dropout, multiply the current value of the $n-1$-th hidden layer h_t^{n-1} and the lagged value of the n-th hidden layer h_{t-1}^n with binary masks $r_{t,h_t^{n-1}}^{(k)} \in \mathbb{R}^{D_{h^{n-1}}}$ and $r_{t,h_{t-1}^n}^{(k)} \in \mathbb{R}^{D_{h^n}}$, respectively:

$$\underbrace{\overline{h}_t^{n-1}}_{D_{h^{n-1}} \times 1} = \underbrace{r_{t,h_t^{n-1}}^{(k)}}_{D_{h^{n-1}} \times 1} \odot \underbrace{h_t^{n-1}}_{D_{h^{n-1}} \times 1}, \quad r_{t,h_t^{n-1},i}^{(k)} \overset{iid}{\sim} Bernoulli(p_{h_t^{n-1}}), \quad i = 1, \dots, D_{h^{n-1}}$$

$$\underbrace{\overline{h}_{t-1}^{n}}_{D_{h^n} \times 1} = \underbrace{r_{t,h_{t-1}^n}^{(k)}}_{D_{h^n} \times 1} \odot \underbrace{h_{t-1}^{n}}_{D_{h^n} \times 1}, \quad r_{t,h_{t-1}^n,i}^{(k)} \overset{iid}{\sim} Bernoulli(p_{h_{t-1}^n}), \quad i = 1, \dots, D_{h^n}$$

where $t \in B^{(k)}$ and $n = 1, \dots, N$ indexes the hidden layer and it is understood that the 0-th layer is the input vector $h_t^0 \equiv \mathcal{X}_t$. $p_{h_t^{n-1}}, p_{h_{t-1}^n} \in [0,1]$ is the probability

[6] When N_S/N_B is not a whole number, $\lfloor N_S/N_B \rfloor$ of the mini-batches will be 3-D tensors with dimensions (N_B, M, K). One batch will contain leftover sequences and will have dimensions $(N_S \% N_B, M, K)$ where % is the modulus operator.

that time t nodes in the $n-1$-th hidden layer and time $t-1$ nodes in the n-th hidden layer are retained, respectively.

(b) *Stochastic Gradient.* Average the gradient over observations in the mini-batch

$$\nabla L(\boldsymbol{\beta}_{j,h}^{(k-1)}, \mathcal{X}^{B^{(k)}}, \boldsymbol{\lambda}^{\mathrm{LSTM}}) = \frac{1}{M} \sum_{t \in B^{(k)}} \nabla L(\boldsymbol{\beta}_{j,h}^{(k-1)}, \mathcal{X}_t, \boldsymbol{\lambda}^{\mathrm{LSTM}})$$

where $\nabla L(\boldsymbol{\beta}_{j,h}^{(k-1)}, \mathcal{X}_t, \boldsymbol{\lambda}^{\mathrm{LSTM}})$ is the gradient of the loss function with respect to the parameters $\boldsymbol{\beta}_{j,h}^{(k-1)}$, evaluated at the time t observation $\mathcal{X}_t$ after applying dropout.

(c) *Learning rate shrinkage.* Update the parameters to $\boldsymbol{\beta}_{j,h}^{(k)}$ using the Adaptive Moment Estimation (Adam) algorithm. The method uses the first and second moments of the gradients to shrink the overall learning rate to zero as the gradient approaches zero.

$$\boldsymbol{\beta}_{j,h}^{(k)} = \boldsymbol{\beta}_{j,h}^{(k-1)} - \gamma \frac{m^{(k)}}{\sqrt{v^{(k)}} + \varepsilon}$$

where $m^{(k)}$ and $v^{(k)}$ are weighted averages of first two moments of past gradients:

$$m^{(k)} = \frac{1}{1-\pi_1^k}(\pi_1 m^{(k-1)} + (1-\pi_1)\nabla L(\boldsymbol{\beta}_{j,h}^{(k-1)}, \mathcal{X}^{B^{(k)}}, \boldsymbol{\lambda}^{\mathrm{LSTM}}))$$

$$v^{(k)} = \frac{1}{1-\pi_2^k}(\pi_2 v^{(k-1)} + (1-\pi_2)\nabla L(\boldsymbol{\beta}_{j,h}^{(k-1)}, \mathcal{X}^{B^{(k)}}, \boldsymbol{\lambda}^{\mathrm{LSTM}})^2)$$

π^k denotes the k-the power of $\pi \in (0,1)$, and $/$, $\sqrt{\cdot}$, and $(\cdot)^2$ are applied elementwise. The default values of the hyperparameters are $m^{(0)} = v^{(0)} = 0$ (initial moment vectors), $\gamma = 0.001$ (initial learning rate), $(\pi_1, \pi_2) = (0.9, 0.999)$ (decay rates), and $\varepsilon = 10^{-7}$ (prevent zero denominators).

(d) *Stopping Critera.* Stop iterating and return $\boldsymbol{\beta}_{j,h}^{(k)}$ if one of the following holds:

- *Early stopping.* At each iteration, use the updated $\boldsymbol{\beta}_{j,h}^{(k)}$ to calculate the loss from the validation sample. Stop when the validation loss has not improved for S steps, where S is a "patience" hyperparameter. By updating the parameters for fewer iterations, early stopping shrinks the final parameters $\boldsymbol{\beta}_{j,h}$ towards the initial guess $\boldsymbol{\beta}_{j,h}^{(0)}$, and at a lower computational cost than ℓ_2 regularization.
- *Maximum number of epochs.* Stop if the number of iterations reaches the maximum number of epochs E. An epoch happens when the full set of the

training sample has been used to update the parameters. If the training sample has T_E observations and each mini-batch has M observations, then each epoch would contain $\lceil T_E/M \rceil$ iterations (after rounding up as needed). So the maximum number of iterations is bounded by $E \times \lceil T_E/M \rceil$.

4. *Ensemble forecasts.* Repeat steps 1. and 2. over different random seeds R and save each of the estimated parameters $\widehat{\boldsymbol{\beta}}^{LSTM}_{j,h,T_E}(\mathcal{X}^t, \boldsymbol{\lambda}^{\text{LSTM}}, R)$. Then construct the out-of-sample forecast as the average across all 20 resulting forecasts. Ensemble can be considered as a regularization method because it aims to guard against overfitting by shrinking the forecasts toward the average across different random seeds. The random seed affects the random draws of the parameter's initial starting value $\boldsymbol{\beta}^{(0)}_{j,h}$, the sequences selected in each mini-batch $B^{(k)}$, and the dropout mask $r^{(k)}_t$.

Regularization

To prevent overfitting and improve generalization, we incorporate several forms of regularization into the machine learning algorithm:

- *L_1 and L_2 penalties.* The loss function includes both an L_1 (lasso) penalty, which encourages sparsity, and an L_2 (ridge) penalty, which shrinks weights toward zero. These penalties are applied to all model parameters and selected via cross-validation. See the loss function in the preceding section for the exact formulation.

- *Dropout.* Dropout is implemented within the LSTM network on both input and recurrent nodes. At each training step, a random subset of nodes is deactivated, encouraging the network to rely on different subsets of units. This functions similarly to an L_1 penalty and promotes sparsity and robustness. See the LSTM section above for implementation details.

- *Early stopping.* Training halts when the validation loss fails to improve for a fixed number of iterations, shrinking parameter estimates toward their initial values and helping prevent overfitting.

- *Ensemble forecasts.* Forecasts are constructed by averaging across 20 different random seeds, each corresponding to different initial weights. This averaging reduces forecast variance and mitigates sensitivity to any single realization of the training process.

Hyperparameters

Let $\boldsymbol{\lambda}^{\text{LSTM}} \equiv [\lambda_1, \lambda_2, \gamma, \pi_1, \pi_2, p, N, (D_{h^n})_{n=1}^{N}, M, E, S]'$ collect all the hyper-parameters that control the LSTM network's complexity and prevent the model from overfitting the data. The number of hidden layers N and the number of neurons $D_{h^1}, \ldots, D_{h^N}$ in each hidden layer are hyper-parameters that characterize the network's architecture. To choose the number of neurons in each layer, we apply a geometric pyramid rule where the dimension of each additional hidden layer is half that of the previous hidden layer. We select the best LSTM architecture iteratively by minimizing the pseudo out-of-sample mean-squared error from rolling forecasts over the validation sample.

Table A.5 reports the hyper-parameters for the LSTM network and its estimation. Hyper-parameters reported as a range or a set of values are cross-validated. The hyper-parameters are estimated by minimizing the mean-square loss over pseudo out-of-sample forecast errors generated from rolling regressions through the validation sample. The pseudo out-of-sample forecasts are ensemble averages implied by parameters based on different random seeds R.

While the scale of the L_1 and L_2 penalty values used in our LSTM models may appear small in magnitude, ranging from 10^{-6} to 10^{-2}, they still play an important role in the regularization. In neural networks, penalties are typically applied to large numbers of weights, so even small penalty values can accumulate meaningfully across the total loss function. Moreover, these penalties can interact with other regularization methods such as dropout and early stopping, making larger penalty values unnecessary. Similar penalty magnitudes are also used in related work, such as Gu et al. (2020), who apply L_1 penalties in the range of 10^{-5} to 10^{-3} asset pricing applications.

For predicting stock returns, we set the L_1 penalty parameter to zero, removing L_1 regularization from the loss function. Given the typically low signal-to-noise ratio in stock returns, this reduces the risk of underfitting caused by excessive sparsity, especially near critical turning points such as crisis periods. In deep neural networks such as LSTMs, L_1 penalties can eliminate important nonlinear interactions and temporal dependencies needed to detect turning points. Our validation experiments confirmed that models with active L_1 penalties consistently underperformed by losing the ability to capture major crises. Instead, we rely on L_2 penalties, dropout, and early stopping, which regularize without forcing weights to zero, preserving the model's sensitivity to turning point dynamics.

Adaptive LSTM Architecture Selection We allow the LSTM architecture to evolve over time using a simple, adaptive updating procedure. At each period in the testing sample, the machine selects the architecture (number of hidden layers and neurons per layer) that minimized out-of-sample forecast errors in the preceding period. The candidate architectures considered span various combinations of hidden layers and neurons per layer, as listed in Table A.5. The architecture is updated quarterly by using the forecast performance from the most recent quarter. This systematic approach allows the machine to adjust its specification over time based on evolving patterns in the data, while avoiding look-ahead bias or overfitting to future outcomes.

LSTM Parameter Counts The number of parameters in an LSTM network can be calculated using the general formula: for layer D_l with input dimension D_d and hidden dimension D_h, the parameter count is $4D_h(D_d + D_h + D_1)$, accounting for the four gates (input, forget, output, cell) each with weight matrices connecting inputs and previous hidden states plus bias vectors. For stock return forecasting with 135 input variables, the candidate architectures range from 2,240 parameters (single-layer, 4 neurons) to 18,176 parameters (5-layer, 16 neurons per layer). For earnings growth prediction with 135 input variables, the architectures span from 16,064 parameters (4-layer, 16 neurons per layer) to 183,296 parameters (5-layer, 64 neurons per layer).

Machine Variables to Be forecast

Returns and price growth When evaluating the MSE ratio of the machine relative to that of a benchmark survey, we use the machine forecast for the return or price growth measure that most closely corresponds to the concept that survey respondents are asked to predict:

1. CFO survey asks respondents about their expectations for the S&P 500 return over the next 12 months. Following Nagel and Xu (2022), we interpret the survey to be asking about $r^d_{t,t+12}$, the one-year CRSP value-weighted return (including dividends) from the current survey month to the same month one year ahead.

2. Gallup/UBS survey respondents report the return (including dividends) they expect on their own portfolio one year ahead. We interpret the survey to be asking about $r^d_{t,t+12}$, the one-year CRSP value-weighted return(including dividends) from the current survey month to the same month one year ahead.

3. Livingston survey respondents provide 12-month ahead forecasts of the S&P 500 index. We convert the level forecast to price growth forecast by taking the log difference between the 12-month ahead level forecast and the nowcast of the S&P 500 index for the current survey month. Therefore, we interpret the survey to be asking about the one-year price growth in the S&P 500 index.

Table A.5: Candidate hyper-parameters for the machine learning forecast

Variable Horizon	Earnings Growth 1-Year	Earnings Growth 4-5 LTG	Earnings Growth 1-10 LTG	Stock Returns 1-Year	Price Growth 1-Year
(a) Elastic Net					
L_1 penalty λ_1	$[10^{-2}, 10^1]$	$[10^{-2}, 10^1]$	$[10^{-2}, 10^1]$	$[10^{-4}, 10^1]$	$[10^{-4}, 10^1]$
L_2 penalty λ_2	$[10^{-2}, 10^1]$	$[10^{-2}, 10^1]$	$[10^{-2}, 10^1]$	$[10^{-4}, 10^1]$	$[10^{-4}, 10^1]$
Training window T_E	$4, 6, 8, 10$	$4, 6, 8, 10, 12$	$4, 6, 8, 10, 12$	$4, 5, 6, 7$	$4, 5, 6, 7$
Validation window T_V	$4, 6, 8, 10$	$4, 6, 8, 10, 12$	$4, 6, 8, 10, 12$	$4, 5, 6, 7$	$4, 5, 6, 7$
(b) Long Short-Term Memory Network					
L_1 penalty λ_1	$[10^{-6}, 10^{-2}]$	$[10^{-6}, 10^{-2}]$	$[10^{-6}, 10^{-2}]$	$[0.0]$	$[10^{-6}, 10^{-2}]$
L_2 penalty λ_2	$[10^{-6}, 10^{-2}]$	$[10^{-6}, 10^{-2}]$	$[10^{-6}, 10^{-2}]$	$[10^{-4}, 10^{-2}]$	$[10^{-6}, 10^{-2}]$
Learning rate γ	0.001	0.001	0.001	0.001	0.001
Gradient decay π_1, π_2	$0.9, 0.999$	$0.9, 0.999$	$0.9, 0.999$	$0.9, 0.999$	$0.9, 0.999$
Dropout input p_x	0.8	0.8	0.8	0.8	0.8
Dropout recurrent p_h	0.8	0.8	0.8	0.4	0.5
Hidden layers N	$1, 3, 5$	$1, 3, 5$	$1, 3, 5$	$1, 3, 5$	$1, 3, 5$
Neurons per layer	$16, 32, 64$	$16, 32, 64$	$16, 32, 64$	16	$4, 8, 16$
Mini-batch size M	4	4	4	4	4
Max epochs E	$10,000$	$10,000$	$10,000$	$10,000$	$10,000$
Early stopping S	20	20	20	80	20
Random seeds R	$1, \ldots, 20$	$1, \ldots, 20$	$1, \ldots, 20$	$1, \ldots, 20$	$1, \ldots, 20$
Training window T_E	$4, 8, 12$	$3, 7, 12$	$3, 7, 12$	$5, 7, 30$	$5, 7$
Validation window T_V	$4, 8, 12$	$3, 7, 12, 20$	$3, 7, 12, 20$	$3, 4$	$5, 7, 20$

Notes: This table reports the hyperparameters considered in the machine learning algorithm for each estimator.

4. Bloomberg Consensus Forecasts asks survey respondents about the end-of-year closing value of the S&P 500 index. We interpret the survey to be asking about the h-month price growth in the S&P 500 index. The horizon of the forecast changes depending on when in the year the panelists are answering the survey.

5. Michigan Survey of Consumers (SOC) asks respondents about their perceived probability that an investment in a diversified stock fund would increase in value in the year ahead. We interpret the question to be asking about the one-year price growth in the S&P 500 index.

6. Conference Board (CB) survey asks respondents about their categorical belief on whether they expect stock prices to increase, decrease, or stay the same over the next year. We interpret the question to be asking about the one-year price growth in the S&P 500 index.

Earnings growth (IBES "Street" Earnings) For earnings growth forecasts, we use a quarterly S&P 500 total earnings series based on IBES street earnings per share (EPS), as described above. Street earnings exclude discontinued operations, extraordinary charges, and other non-operating items, making them better aligned with the earnings measure targeted by survey respondents. We convert EPS to total earnings using the S&P 500 index divisor and use the resulting quarterly series directly, prior to any monthly interpolation, since the machine learning algorithm operates at a quarterly frequency. The IBES street earnings series spans 1983:Q4 to 2021:Q4.

For Long-Term Growth (LTG) forecasts, IBES defines LTG as the "expected annual increase in operating earnings over the company's next full business cycle. These forecasts refer to a period of between three to five years." We compare survey responses of LTG against machine forecasts under alternative interpretations of LTG. First, we consider machine forecasts of annual five-year forward growth, i.e., annual earnings growth from four to five years ahead (Bianchi et al. (2024)). Second, we consider machine forecasts of annualized 5-year growth, i.e., annual earnings growth from current quarter to five years ahead, following the interpretation in Bordalo et al. (2019). Third, we consider machine forecasts of annualized earnings growth from one to 10 years ahead, following the interpretation in Nagel and Xu (2022)

Machine Input Data

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. The predictors below are listed as elements of $y_{j,t}$, $\hat{\mathbf{G}}_{jt}'$, or $\mathbf{W}_{jt}'$ for variables.

Stock return and price growth predictor variables and specifications For y_j equal to CRSP value-weighted returns or S&P 500 price index growth, we first predict the one-year log stock return or price growth that is expected to occur h quarters into the future from time $t + h - 4$ to $t + h$, i.e., $\mathbb{E}_t[r_{t+h-4,t+h}]$. For horizons longer than one year, since the h-quarter long horizon return is the sum of one-year returns between time t to $t + h$, we first forecast the forward one-year returns separately and then add the components together to get machine forecasts of h-quarter long horizon returns. The forecasting model considers the following variables:

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

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. $\mathbf{G}^Q_{D,t-k}$, for $k = 0$ 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 $\mathbf{G}_{D,t}(\mathbf{w})$.

4. *LDA topics* $F_{k,t-j}$, for topic $k = 1, 2, ...50$ and $j = 0, 1$. The value of the topic k at time t is the average weights of topic k of all articles published at t.

5. *Macro data surprises* from the money market survey. The macro news include, GDP growth (Q/Q percentage change), core CPI (Month/Month change), unemployment rate (percentage point), and nonfarm payroll (month/month change). We include first release, second release, and final release for GDP growth. This constitutes six macro data surprises per quarter.

6. *FOMC surprises* are defined as the changes in the current-month, 1, 2, 6, 12, and 24 month-ahead federal funds futures (FFF) contract rate and the changes in the 1, 2, 4, and 8 quarter-ahead Eurodollar (ED) futures contracts, from 10 minutes before to 20 minutes after each FOMC announcement. When benchmarking against a survey, we use the last FOMC meeting before the survey deadline to compute surprises. For surveys that do not have a clear deadline, we compute surprises using from the last FOMC in the first month of the quarter. When benchmarking against moving average, we use the last FOMC meeting before the end of the first month in each quarter to compute surprises. This leaves 10 FOMC surprise variables per quarter.

7. *Stock market jumps* are accumulated 30-minute window negative and positive jumps in the S&P 500 around news events over the previous quarter.

8. *Long-term growth of earnings*: 5-year growth of the SP500 earnings per share.

9. *Short rates.* When forecasting returns or price growth, the machine controls for the current nominal short rate, $\ln(1 + 3MTB_t/100)$, imposing a unit coefficient. This is equivalent to forecasting the future return minus the current short rate.

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.[7] 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.

Earning growth predictor variables and specifications For y_t equal to the log earning growth, we construct a forecasted value for y_t, denoted $\hat{y}_{t|t-h}$, based on information known up to time t using the following variables:

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. $\mathbf{G}^Q_{D,t-k}$, for $k = 0$ 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 $y_{j,t}$ on $\mathbf{G}_{D,t}(\mathbf{w})$.

4. *LDA factors* $F_{k,t-j}$, for topic $k = 1, 2, ...50$ and $j = 0, 1$. The value of the topic k at time t is the average weights of topic k of all articles published at t.

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

5. *Macro data surprises* from the money market survey. The macro news include, GDP growth (Q/Q percentage change), core CPI (Month/Month change), unemployment rate (percentage point), and nonfarm payroll (month/month change). We include first release, second release, and final release for GDP growth. This constitutes six macro data surprises per quarter.

6. *FOMC surprises* are defined as the changes in the current-month, 1, 2, 6, 12, and 24 month-ahead federal funds futures (FFF) contract rate and the changes in the 1, 2, 4, and 8 quarter-ahead Eurodollar (ED) futures contracts, from 10 minutes before to 20 minutes after each FOMC announcement. When benchmarking against a survey, we use the last FOMC meeting before the survey deadline to compute surprises. For surveys that do not have a clear deadline, we compute surprises using from the last FOMC in the first month of the quarter. When benchmarking against moving average, we use the last FOMC meeting before the end of the first month in each quarter to compute surprises. This leaves 10 FOMC surprise variables per quarter.

7. *Stock market jumps* are accumulated 30-minute window negative and positive jumps in the S&P 500 around news events over the previous quarter.

Table A.6: Number of Right-Hand Side Variables

	Stock Return	Earnings
Macro Factors	10 (0-1 lag)	10 (0-1 lag)
Financial Factors	10 (0-1 lag)	10 (0-1 lag)
Daily Factors	10 (0-1 lag)	10 (0-1 lag)
LDA Media-Sentiment Factors	50	50
FOMC Surprises	10	10
Macro Data Surprises	6 (0-1 lag)	6 (0-1 lag)
Other Predictors	0	3
Total	132	135

Note: This table shows the number of predictors included for each forecast target.

Forecast Decomposition Details

Let$\mathcal{X}_t$ denote the vector of inputs available as of time t. Let $\mathcal{X}^t$ denote the history of inputs from $1, ..., t$. The LSTM function estimated at each t can be arbitrarily well approximated as the sum of three components:

$$G(\cdot) \approx \underbrace{b_{y,t}}_{\text{Local mean}} + \underbrace{h_t(h_{t-1}, \bar{\mathcal{X}}_t; \vec{b}_t, \widehat{\boldsymbol{\beta}}_t)}_{\text{Nonlinear history}} + \underbrace{\nabla G(\cdot)' \mathcal{X}_t}_{\text{Current information}} . \tag{A.10}$$

The first b_y term is an outer bias term. Since we estimate this over time, this evolves slowly and is akin to a local mean. The second, $h_t(\cdot)$, term is the same as the LSTM function $W^{(yh^N)} h_t^N(\cdot)$ given in subsection 3.4, a function of lagged hidden layers $h_{t-1}(\cdot)$ (functions of $\mathcal{X}^{t-1}$), current estimates of the inner bias terms $\vec{b}_t \equiv (b_{c^n}, b_{f^n}, b_{i^n}, b_{o^n})'$ and estimated weights $\widehat{\boldsymbol{\beta}}_t$, but holding fixed current inputs $\mathcal{X}_t$ at the sample mean over the training sample (denoted as fixed by $\bar{\mathcal{X}}_t$). Since each input variable in $\mathcal{X}_t$ is standardized before being passed to the LSTM estimator, this means that the initial hidden layer $h_t^0 \equiv \mathcal{X}_t$ is replaced a vector of zeros. The third term captures the impact of current inputs $\mathcal{X}_t$ with a linear approximation of their role that holds fixed the history $\mathcal{X}^{t-1}$. Here, $\nabla G(\cdot)$ denotes the *integrated gradient* of the LSTM function with respect to the information variables$\mathcal{X}_t$, where the gradient is with respect to the ith input variable. The gradient $\nabla G(\cdot) \in \mathbb{R}^D$, where D is the dimension of the input vector $\mathcal{X}_t$, is defined as

$$\nabla G(\cdot)_i \equiv \int_{\delta=0}^{\delta=1} \frac{\partial G(\delta \mathcal{X}_t; \cdot)}{\partial \mathcal{X}_{i,t}} d\delta \approx \frac{1}{M} \sum_{m=0}^{M} \frac{\partial G(\frac{m}{M} \mathcal{X}_t; \cdot)}{\partial \mathcal{X}_{i,t}}, \quad i = 1, \ldots, D. \tag{A.11}$$

In practice, we use a discrete approximation with $M = 300$ equally spaced points, which provides a very accurate approximation of our LSTM functions; see Figure A.1 below. Note that nonlinearities in the current information are captured by the gradient magnitudes.

The above approximation is used for decomposing machine forecasts over time. For decomposing survey respondent forecasts, we use machine learning to learn about the forecasting models of the survey respondents. Specifically, we ask the machine to predict the survey respondent's forecast out-of-sample, by minimizing the distance between the survey response at time t and using the history of input data up to t. To mitigate overfitting, we estimate recursive out-of-sample forecasts where the loss function is penalized using ridge and lasso penalties. For example, for fitting 1-year (4 quarter) ahead investor forecasts the penalty parameters λ_1 and λ_2 are given by:

$$\widehat{\boldsymbol{\lambda}} = \underset{\boldsymbol{\lambda}}{argmin}\left\{\frac{1}{4}\sum_{h=1}^{4}\left[\frac{1}{T_V}\sum_{\tau=1}^{T_V}\left(\mathbb{F}_{\tau+h}[y_{j,\tau+h+4}] - G(\cdot)\right)^2\right]\right. \tag{A.12}$$

$$\left. + \underbrace{\lambda_1 \sum_{h=1}^{4}\sum_{k=1}^{K}\left|\widehat{\boldsymbol{\beta}}_{\tau,k}\right|}_{L_1 \text{ Penalty}} + \underbrace{\lambda_2 \sum_{h=1}^{4}\sum_{k=1}^{K}\left(\widehat{\boldsymbol{\beta}}_{\tau,k}\right)^2}_{L_2 \text{ Penalty}}\right\} \tag{A.13}$$

In the loss function, $h \in \{1, 2, 3, 4\}$ indexes the forecast horizons for the predictive model. $\mathbb{F}_{\tau+h}[y_{j,\tau+h+4}]$ denotes the investor forecast made at time $\tau + h$ for the value of variable y_j four quarters ahead, i.e., for period $\tau + h + 4$. In other words, at date $\tau + h$, survey respondents report their expectation of y_j one year into the future. T_V denotes the length of the validation sample periods. The function $G(\cdot)$ represents the LSTM forecast at horizon h, and $\widehat{\boldsymbol{\beta}}_\tau$ are the corresponding network parameters estimated using training data and shared regularization penalties λ_1 and λ_2. All other LSTM-related hyper-parameters match those used in the machine forecast. After choosing the hyperparameters as above, we construct a final true out of sample prediction of the survey forecast one quarter ahead of the survey response date. The final fitted survey forecast takes the form

$$\hat{\mathbb{F}}(\cdot) \approx \underbrace{b_{y,t}}_{\text{Local mean}} + \underbrace{h_t(h_{t-1}, \bar{\mathcal{X}}_t; \vec{b}_t, \widehat{\boldsymbol{\beta}}_t)}_{\text{Nonlinear history}} + \underbrace{\nabla G(\cdot)'\mathcal{X}_t}_{\text{Current information}} + \underbrace{u_t}_{\text{Residual}}$$

where the residual captures the error in the fitted model and and represents information intangible to the machine such as private information or forecaster judgment.

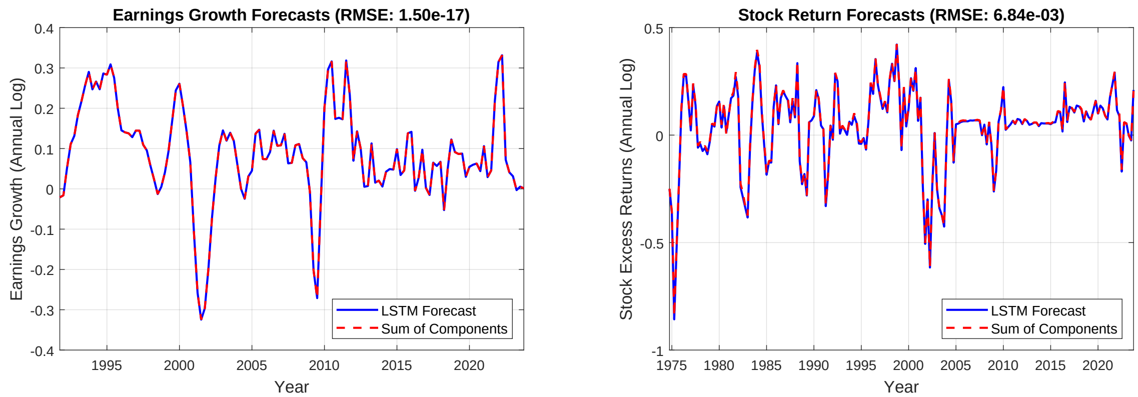
Categorizing Data Inputs

This section documents the grouping of machine input variables. Dynamic and LDA factors are given rough names using the procedure described in following subsection.

Economic Names of Factors

Macro, Financial, Daily 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

Figure A.1: Accuracy of Forecast Decomposition



Notes: This figure validates the accuracy of the integrated gradients decomposition by comparing the original LSTM stock return forecasts (blue solid line) with the reconstructed forecasts obtained by summing all decomposition components (red dashed line). The Root Mean Square Error (RMSE) between the original and reconstructed forecasts is reported in the title. The x-axis indicates the forecasted period with realized outcomes. Stock excess returns are expressed as log annual returns in excess of the 1-year T-bill rate, and earnings growth is measured as the log annual growth rate. NBER recessions are shaded in gray. The sample spans: 1991:Q1-2004:Q4 (training earnings), 1973:Q1-2004:Q4 (training stock returns), 2005:Q1-2023:Q4 (testing).

output or unemployment. Following Ludvigson and Ng (2007), 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 $\tilde{x}_{it}$ 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 $\tilde{x}_{it}$ 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 $\tilde{x}_{it}$ 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 $\tilde{x}_{it}$. 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.

LDA Factors We follow Bybee et al. (2021) in assigning economic names to the Latent Dirichlet Allocation (LDA) factors. The kth LDA factor $F_{k,t}$ at period t is defined as the average attention weight $\theta_{i,k,t}$ allocated to topic k across all articles published during the period. A topic is a probability distribution over words. Formally, the kth topic is a V-dimensional vector $\phi_{k,t}$ in the kth row of the topic-word distribution $\Phi_t = [\phi_{1,t}, \ldots, \phi_{K,t}]'$, where K is the total number of topics and V is the number of unique words in the corpus. Since the parameters $\theta_{i,k,t}$ and $\phi_{k,t}$ are estimated recursively over real-time quarterly vintages of news articles, the estimates may vary over the training samples and are likewise denoted with a t subscript.

To summarize the economic content of the topic that prevailed the most consistently across time, we select the key word that is expected to occur with the highest average

Table A.7: Predictor Variable Grouping for Machine Learning Forecasts

Category	Factor/Variable	Label
(a) Fundamentals		
Macro Factor: Employment	$G_{1,M,t}$, $G_{2,M,t}$, $G_{3,M,t}$	Avg hourly earnings growth, Help-wanted index, Avg weekly hours
Macro Factor: Prices	$G_{4,M,t}$, $G_{5,M,t}$, $G_{7,M,t}$	Price index - pers. consumption goods, Price index - nondurable
Macro Factor: Output & Income	$G_{6,M,t}$, $G_{8,M,t}$	Personal interest income, Energy production index
Macro Factor: Consumption	$G_{9,M,t}$, $G_{10,M,t}$	Real pers. consumption - nondurable, Real pers. consumption - services
Daily Factor: Treasuries	$G_{1,D,t}$, $G_{4,D,t}$, $G_{5,D,t}$, $G_{7,D,t}$, $G_{8,D,t}$, $G_{10,D,t}$	10Y-Fed spread, 6M-3M spread, 10Y-3M spread
Corporate earnings	Long-term growth	5yr S&P earnings
Short-term interest rates	Short rates	3M T-bill
FOMC surprises	FOMC surprises	Fed funds fut (curr, 1-24m), Eurodollar fut (1-8q)
Macro data surprises	Macro surprises	GDP, Core CPI, Unemp., Payrolls
Lagged earnings growth†	Lagged dependent var	1-2q lags
(b) Non-fundamentals		
Financial Factor: Size/BM Portfolios	$G_{1,F,t}$	Size/BM portfolio high
Financial Factor: Equity Risk	$G_{2,F,t}$, $G_{3,F,t}$	High minus low
Financial Factor: Price/Yld/Div	$G_{4,F,t}$, $G_{5,F,t}$, $G_{7,F,t}$, $G_{8,F,t}$	Dividend reinvestment return, Dividend-price ratio
Financial Factor: Industries	$G_{6,F,t}$, $G_{9,F,t}$, $G_{10,F,t}$	Restaurants/hotels/motels, Textiles
Daily Factor: Corporate Risk	$G_{2,D,t}$, $G_{3,D,t}$, $G_{6,D,t}$, $G_{9,D,t}$	Eurodollar-Fed spread, A-rated corp minus 10Y, Commercial paper spreads
Stock Market Jumps	S&P 500 jumps	Positive/negative jumps
Lagged stock returns‡	Lagged dependent var	1-2q lags
Lagged survey forecast§	Lagged Long-Term Growth (LTG)	1-2q lags
(c) LDA Media-Sentiment		
Media-Sentiment: Financial mkt	$LDA_{2,t}$, $LDA_{4,t}$, $LDA_{5,t}$, $LDA_{7,t}$, $LDA_{8,t}$, $LDA_{9,t}$, $LDA_{11,t}$, $LDA_{13,t}$, $LDA_{20,t}$, $LDA_{24,t}$, $LDA_{26,t}$, $LDA_{27,t}$, $LDA_{31,t}$, $LDA_{34,t}$, $LDA_{35,t}$, $LDA_{37,t}$, $LDA_{39,t}$, $LDA_{42,t}$, $LDA_{44,t}$, $LDA_{50,t}$	Intl exch., Options/VIX, FX/metals, Exchanges, Trading, Payouts, IPOs
Media-Sentiment: Banking	$LDA_{15,t}$, $LDA_{18,t}$, $LDA_{28,t}$, $LDA_{33,t}$, $LDA_{36,t}$, $LDA_{43,t}$, $LDA_{30,t}$, $LDA_{40,t}$	Mortgages, NPLs
Media-Sentiment: Asset mgmt	$LDA_{6,t}$, $LDA_{10,t}$, $LDA_{41,t}$	Mutual funds
Media-Sentiment: Political/Govt	$LDA_{1,t}$, $LDA_{16,t}$, $LDA_{32,t}$, $LDA_{46,t}$	Clintons, Public/private, Watchdogs
Media-Sentiment: Industry	$LDA_{14,t}$, $LDA_{17,t}$, $LDA_{21,t}$, $LDA_{19,t}$, $LDA_{23,t}$, $LDA_{3,t}$, $LDA_{45,t}$	Auto, Airlines, Couriers, Chemicals
Media-Sentiment: Intl affairs	$LDA_{48,t}$, $LDA_{47,t}$	UK, Nuclear/North Korea
Media-Sentiment: Other	$LDA_{12,t}$, $LDA_{29,t}$, $LDA_{38,t}$, $LDA_{49,t}$	Futures/indices

Note: †Lagged earnings growth grouped as fundamental (earnings forecasts only). ‡Lagged stock returns grouped as non-fundamental (return forecasts only). §IBES survey forecast included in earnings specifications only. Abbreviations: Yld = yield, div = dividend, fut = futures, exch = exchanges, FX = foreign exchange, NPLs = nonperforming loans, Auto = automotive.

probability across our testing sub-sample of interest, i.e., the largest element in the V-dimensional vector $\frac{1}{T}\sum_{t=1}^{T}\phi_{k,t}$ where t indexes the time periods of the evaluation sample of length T. We use Table 6 of Bybee et al. (2021) to map the top key word for each topic to their topic label. The authors have manually assigned a label to each topic based on their reading of the key terms list. We also use the same table to categorize each topic label into broader meta-topics. For example, the key word "clinton" has the highest average probability of occurring under topic $k = 1$ across an testing sub-sample of 2005:Q1 to 2023:Q4. Therefore,

Table A.8: Economic Interpretation of the Factors

Factor	Series with Largest R^2	Label	R^2
Macro Factors			
$G_{1,M,t}$	Average Hourly Earnings Growth	Macro Factor: Employment	0.447
$G_{2,M,t}$	Help-Wanted Index	Macro Factor: Employment	0.253
$G_{3,M,t}$	Average Weekly Hours	Macro Factor: Employment	0.317
$G_{4,M,t}$	Price Index - Personal Consumption Goods	Macro Factor: Prices	0.154
$G_{5,M,t}$	Price Index - Personal Consumption Goods	Macro Factor: Prices	0.400
$G_{6,M,t}$	Personal Interest Income	Macro Factor: Output & Income	0.229
$G_{7,M,t}$	Price Index - Personal Consumption Non-durable	Macro Factor: Prices	0.084
$G_{8,M,t}$	Energy Production Index	Macro Factor: Output & Income	0.175
$G_{9,M,t}$	Real Personal Consumption - Nondurable Private	Macro Factor: Consumption	0.112
$G_{10,M,t}$	Real Personal Consumption - Services Total	Macro Factor: Consumption	0.226
Financial Factors			
$G_{1,F,t}$	Size/BM Portfolio High	Financial Factor: Size/BM Portfolios	0.131
$G_{2,F,t}$	High Minus Low	Financial Factor: Equity Risk Factors	0.027
$G_{3,F,t}$	High Minus Low	Financial Factor: Equity Risk Factors	0.100
$G_{4,F,t}$	Dividend Reinvestment Return	Financial Factor: Prices, Yields, Dividends	0.035
$G_{5,F,t}$	Dividend-Price Ratio	Financial Factor: Prices, Yields, Dividends	0.121
$G_{6,F,t}$	Restaurants, Hotels, Motels Industry	Financial Factor: Industries	0.058
$G_{7,F,t}$	Dividend-Price Ratio	Financial Factor: Prices, Yields, Dividends	0.055
$G_{8,F,t}$	Dividend-Price Ratio	Financial Factor: Prices, Yields, Dividends	0.035
$G_{9,F,t}$	Textiles Industry	Financial Factor: Industries	0.042
$G_{10,F,t}$	Textiles Industry	Financial Factor: Industries	0.039
Daily Factors			
$G_{1,D,t}$	10-year treasury bond yield minus Fed Funds	Daily Factor: Treasuries	0.221
$G_{2,D,t}$	3-Month Eurodollar Deposits minus Fed Funds	Daily Factor: Corporate Risk	0.093
$G_{3,D,t}$	A Rated Corp. Bond Yield minus Y10-Tbond	Daily Factor: Corporate Risk	0.566
$G_{4,D,t}$	6-month minus 3-month treasury bill yield	Daily Factor: Treasuries	0.276
$G_{5,D,t}$	6-month minus 3-month treasury bill yield	Daily Factor: Treasuries	0.418
$G_{6,D,t}$	1-M A2/P2/F2 minus Aa Fin. Commerical Paper	Daily Factor: Corporate Risk	0.203
$G_{7,D,t}$	10-year minus 3-month treasury bond yield	Daily Factor: Treasuries	0.341
$G_{8,D,t}$	10-year minus 3-month treasury bond yield	Daily Factor: Treasuries	0.409
$G_{9,D,t}$	A Rated Corp. Bond Yield minus Y10-Tbond	Daily Factor: Corporate Risk	0.296
$G_{10,D,t}$	6-month minus 3-month treasury bill yield	Daily Factor: Treasuries	0.635

Note: This table reports the series with the largest marginal R^2 for the factor specified in the first column. The marginal R^2 is computed from regressions of each individual series onto the corresponding factor.

we label $F_{1,t}$ according to the label for which "clinton" belongs to, which is the topic label "Clintons" that falls under a broader meta-topic label "Political Leaders."

Table A.9: Economic Interpretation of LDA Factors

Factor	Meta-Topic	Topic	**Factor**	Meta-Topic	Topic
$LDA_{1,t}$	Politics	Clintons	$LDA_{26,t}$	Fin Mkts	Trading
$LDA_{2,t}$	Fin Mkts	Intl exchanges	$LDA_{27,t}$	Fin Mkts	Trading
$LDA_{3,t}$	Industry	Couriers	$LDA_{28,t}$	Banks	Mortgages
$LDA_{4,t}$	Fin Mkts	Options/VIX	$LDA_{29,t}$	Activism	Futures/indices
$LDA_{5,t}$	Fin Mkts	FX/metals	$LDA_{30,t}$	Banks	NPLs
$LDA_{6,t}$	Asset Mgrs	Mutual funds	$LDA_{31,t}$	Fin Mkts	Payouts
$LDA_{7,t}$	Fin Mkts	Exchanges	$LDA_{32,t}$	Govt	Public/private
$LDA_{8,t}$	Fin Mkts	FX/metals	$LDA_{33,t}$	Banks	Mortgages
$LDA_{9,t}$	Fin Mkts	Intl exchanges	$LDA_{34,t}$	Fin Mkts	Exchanges
$LDA_{10,t}$	Asset Mgrs	Mutual funds	$LDA_{35,t}$	Fin Mkts	IPOs
$LDA_{11,t}$	Fin Mkts	Trading	$LDA_{36,t}$	Banks	Mortgages
$LDA_{12,t}$	Activism	Futures/indices	$LDA_{37,t}$	Fin Mkts	Trading
$LDA_{13,t}$	Fin Mkts	Trading	$LDA_{38,t}$	Activism	Futures/indices
$LDA_{14,t}$	Transport	Automotive	$LDA_{39,t}$	Fin Mkts	Exchanges
$LDA_{15,t}$	Banks	Mortgages	$LDA_{40,t}$	Banks	NPLs
$LDA_{16,t}$	Govt	Public/private	$LDA_{41,t}$	Asset Mgrs	Mutual funds
$LDA_{17,t}$	Transport	Automotive	$LDA_{42,t}$	Fin Mkts	Trading
$LDA_{18,t}$	Banks	Mortgages	$LDA_{43,t}$	Banks	Mortgages
$LDA_{19,t}$	Transport	Airlines	$LDA_{44,t}$	Fin Mkts	Trading
$LDA_{20,t}$	Fin Mkts	Trading	$LDA_{45,t}$	Industry	Chemicals/paper
$LDA_{21,t}$	Transport	Automotive	$LDA_{46,t}$	Govt	Watchdogs
$LDA_{22,t}$	Fin Mkts	Exchanges	$LDA_{47,t}$	Mideast/Terror	Nuclear/NK
$LDA_{23,t}$	Transport	Airlines	$LDA_{48,t}$	Intl Affairs	UK
$LDA_{24,t}$	Fin Mkts	Trading	$LDA_{49,t}$	Activism	Futures/indices
$LDA_{25,t}$	Fin Mkts	Options/VIX	$LDA_{50,t}$	Fin Mkts	Exchanges

Notes: This table summarizes the economic interpretation of each LDA factor. Meta-topic and topic labels are based on keyword distributions following Bybee et al. (2021). Abbreviations: Fin Mkts = Financial Markets, Asset Mgrs = Asset Managers/I-Banks, FX = Currencies/Metals, NPLs = Nonperforming Loans, NK = North Korea, Govt = Government.

Economic Interpretation of Nonlinear History

Given that the nonlinear history of inputs plays an important role in the machine forecasts, it is reasonable to ask what economic interpretation we can give to these components. We caution that any labeling of these components is imperfect, because they are influenced to some degree by all the variables in our input dataset and because the relations among them are nonlinear. Nonetheless, it is useful to show that the component captures relevant macroeconomic information. Table A.10 shows the marginal R^2 of the nonlinear history component for the five most relevant variables. The marginal R^2 is the R^2 statistic from regressions of each of the inputs onto the nonlinear history, using the full sample of data.

What we observe is that this component is most highly correlated with measures of corporate risk and Treasury returns, for both machine expectations of returns and earnings growth. This suggests that such variables play an important role in predicting turning points.

Table A.10: Top 3 Variables by Marginal R^2 on LSTM Nonlinear History

Lag	Variable	Group	Description	R^2
		Earnings Growth		
Lag 1	fff3	FOMC Surprise	3-Month Ahead Fed Funds Surprise	0.196
	10YTB_FF	Daily Factor: Treasuries	10-year treasury bond yield at constant maturity (% P.A.) minus Fed Funds	0.175
	TED	Daily Factor: Corporate Risk	3-Month Tbill minus 3-Month London Interbank Offered Rate (%)	0.164
Lag 2	10YTB_FF	Daily Factor: Treasuries	10-year treasury bond yield at constant maturity (% P.A.) minus Fed Funds	0.216
	TED	Daily Factor: Corporate Risk	3-Month Tbill minus 3-Month London Interbank Offered Rate (%)	0.209
	6MTB_FF	Daily Factor: Treasuries	6-month treasury bill market bid yield at constant maturity (% P.A.) minus Fed Funds	0.142
		Stock Returns		
Lag 1	MAaa_10YTB	Daily Factor: Corporate Risk	Moody Seasoned Aaa Corporate Bond Yield (% P.A.) minus Y10-Tbond	0.149
	LIBOR	Daily Factor: Corporate Risk	Overnight London Interbank Offered Rate (%)	0.108
	1YTB_3MTB	Daily Factor: Treasuries	1-year treasury bill yield at constant maturity (% P.A.) minus 3M-Tbills	0.094
Lag 2	MAaa_10YTB	Daily Factor: Corporate Risk	Moody Seasoned Aaa Corporate Bond Yield (% P.A.) minus Y10-Tbond	0.145
	MBaa_10YTB	Daily Factor: Corporate Risk	Moody Seasoned Baa Corporate Bond Yield (% P.A.) minus Y10-Tbond	0.060
	jump_plus	Stock Market Jumps	S&P 500 Positive Jump	0.058

Notes: Table reports the top 3 variables ranked by marginal R^2 for regressions of each input variable on the LSTM nonlinear history. Lag 1 and Lag 2 refer to first and second quarterly lags of input variables.