

The Prestakes of Stock Market Investing

FRANCESCO BIANCHI*
Johns Hopkins, CEPR, and NBER

DO LEE†
NYU

SYDNEY C. LUDVIGSON‡
NYU, CEPR, and NBER

SAI MA§
Federal Reserve Board

First Draft: August 2, 2025

This Draft: August 26, 2026

Abstract

We use machine learning to identify and explain “prestakes”—predictable mistakes in financial markets arising from distortions and inefficiencies in participants’ subjective beliefs. The algorithm learns independently, remains attentive to fundamentals, credit risk, and sentiment, and makes abrupt course corrections at critical junctures. The subjective beliefs of investors, professionals, and equity analysts do little of this and instead contain prestakes that are especially prevalent in times of market turbulence. The model achieves higher forecasting accuracy due to its greater complexity and richer data. Market timing strategies that bet against prestakes deliver defensive strategies with large CAPM and Fama-French 5-factor alphas.

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

*Department of Economics, Wyman Park Building, 3100 Wyman Park Drive, Baltimore, MD 21211. (francesco.bianchi@jhu.edu)

†Department of Economics, NYU, 19 W.4th St, 6th Floor, New York, NY 10012. (dql204@nyu.edu)

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

§Federal Reserve Board of Governors, C Ave & 20th Street NW, Washington, DC 20551. (sai.ma@frb.gov)
The views expressed are those of the authors and do not necessarily reflect those of the Federal Reserve Board or the Federal Reserve System. We are grateful to Francesca Bastianello, Lars Peter Hansen, Suproteem Sarkar, and Andrei Shleifer for helpful comments, and to seminar participants at the 2026 NBER Spring Asset Pricing Meeting, 2026 NBER Summer Institute Forecasting and Empirical Methods workshop, the 2026 IAAE Annual Conference, Collegio Carlo Alberto, Harvard Business School, the New York Fed, and the University of Chicago. We thank Xuran Zeng and Luofeng Zhou for excellent research assistance.

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), in which agents efficiently process all available information. Such tests are atheoretical, in the long tradition of forecast-efficiency or orthogonality tests that ask whether forecast errors are predictable from information available to the forecaster, without specifying a structural model of the economy (Mincer and Zarnowitz (1969), Nordhaus (1987), Coibion and Gorodnichenko (2015), Bordalo, Gennaioli, La Porta and Shleifer (2019)).¹ Yet knowing whether a deviation from FIRE 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 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 our explicit measure of non-distorted and efficient expectation formation as a measure of “objective beliefs,” and to any deviation from this measure—whether due to non-rational expectations or informational inefficiency—as a “belief distortion.”²

¹The requirement that forecast errors be orthogonal to all publicly available information is the semi-strong-form market-efficiency criterion of Fama (1970), which in practice coincides with the “full information” benchmark: a forecaster who efficiently uses all publicly available information leaves no predictable component in its errors. We use “FIRE” as the now-standard shorthand for this joint null in the testing literature, not as a structural equilibrium assumption; see the conceptual clarifications below.

²We use “objective” throughout as a compact label for a benchmark constructed to be free of human behavioral bias and to use available information efficiently, in deliberate contrast to the “subjective” beliefs reported in surveys. The term is not a claim to recover the true (physical) probability law governing returns: our benchmark is a feasible statistical estimate that itself takes subjective market data as inputs and, like any forecast, is subject to estimation error and to the unknowability of the true data-generating process.

Our approach begins with the observation that even sophisticated modern-day market participants could have beliefs that are 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, the metric must be rooted in the FIRE ideal, which demands the efficient processing of all available information, but modified for a real-world context. This means accounting for the vast data environment investors actually face, while strictly ensuring all input data were available in real time. Failing to do so—for example, by benchmarking subjective beliefs against low-dimensional prediction models that researchers know 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 human bias and/or cognitive limits to efficient information processing. Third, the measure should be capable of efficiently processing the large amount of information available to real-world investors, 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 volatility over an initial sample that begins in the early 1970s and ends just before the 2004 start of the testing sample, a period that includes several major boom/bust episodes. The algorithm is then tested repeatedly over the remaining sample that ends in 2023. Importantly, the machine is required to do its analysis without

What makes it “objective” is the absence of human cognitive distortion in its construction, not omniscience about the economy.

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.³

The resulting algorithm produces a sequence of machine “beliefs” against which the reported expectations of market participants may be assessed. To quantify predictable errors in real time, the algorithm performs a period-by-period check: it determines if the machine forecast could have identified a mistake in the survey forecast immediately before issuing its own out-of-sample prediction. The magnitude of prestakes can then be quantified as the difference between the survey and machine forecasts whenever predictable mistakes are found, and zero otherwise.

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 of forecasts originating in 2004:Q1-2022:Q4, whose one-year targets are realized in 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 between 21% and 36% 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 61% more accurate than professional equity analysts over the full testing sample.

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 ignore relevant information employed by the machine that could have been used to improve survey forecast accuracy, but evidently was not. It follows that

³In particular, we do not use commercially available large language models, which are known to impart substantial look-ahead bias in forecasting applications (e.g., Levy (2024); Sarkar and Vafa (2024)).

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.

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 or non-linearly enough to evidence of a changing economic landscape. Second, survey respondents show little sign of learning from past mistakes: during the global financial crisis the same input features that drive the machine forecast in one direction drive the IBES survey forecast in the opposite direction, so the machine captures part of the sharp decline in earnings growth and its subsequent recovery while the median equity analyst predicts the rebound a year too early and misses the downturn entirely. Third, survey responses exhibit a form of recency bias, with forecasts largely explained by a local mean in the predicted variable even in times of extreme market turbulence. The machine, by contrast, draws on its long-memory functionality to weigh credit risk variables, fed funds futures market surprises, macroeconomic data, and media-sentiment variables, course-correcting to identify turning points in turbulent times—precisely when it identifies the largest distortions in measured beliefs.

What separates them is selective attention on the survey side: IBES analysts focus excessively on earnings-related information and attend sub-optimally to the broader macroeconomic, financial market, and media signals the machine uses. Selective attention of this kind produces overreaction and underreaction at the same time—analysts overreact to the signals they attend to and underreact to those they do not. Bordalo, Gennaioli, Ma and Shleifer (2020) reconcile this coexistence within a diagnostic expectations version of a dis-

persed information learning model, and our results provide a natural test of their mechanism at a different level of information aggregation: the analyst consensus against the machine, rather than the individual analyst against the consensus.

Meaningful distortions in subjective beliefs raise the possibility that real-world investors have a skewed view of the objective trade-off between risk and reward. This brings us to our second investigation, which compares the risk-return trade-off 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 trade-off in both subjective and objective beliefs, objective risk premia move between 1.9 and 8.1 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 market timing strategies that bet on our measure of objective beliefs earn positive risk-adjusted returns, or “alpha.” Since in practice the machine forecasts differ substantially from survey forecasts only in periods where prestakes are estimated to be large, 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 4.5 and 4.9 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 7.0 and 8.5 percent per annum, respectively. These machine

alphas arrive at an outsized pace during the GFC and the 2022 bear market, but at least half of the cumulative abnormal return is earned during the intervening decade of calm markets, at close to the strategies’ full-sample average rate. By contrast, strategies based on the subjective beliefs of market participants earn essentially zero alphas because they almost never forecast a market return below the Treasury bill rate, let alone losses.

Two conceptual clarifications about our framework are important. First, our approach does not assume that we are measuring the beliefs of an agent in a counterfactual rational expectations model of the world.⁴ Rather, it adopts the FIRE requirement—efficiently processing all available information—from the perspective of an objective forecaster observing an economy where real-world agents may be nonrational. Under this formulation, an objective algorithm (without human bias) that takes as inputs both objective economic data and subjective market data generated by agents with bias can learn to exploit the market’s systematic mistakes to better predict returns, even if they wouldn’t be predictable were agents fully rational (e.g., Nagel and Xu (2022), Bianchi, Ludvigson and Ma (2026)). Second, the algorithm is explicitly designed to capture the bounded rationality and sub-optimal selective attention that arises when investors fail to process all available information efficiently. Our finding that systematic mistakes remain periodically large throughout our sample ending in 2023:Q4—long after advanced data-processing technologies became available to investors—suggests that processing capacity limitations alone cannot fully explain our results.

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—

⁴This distinction matters because, as Lucas (1987) emphasized, rational expectations in the sense of Muth (1961) is a model-consistency axiom that acquires empirical content only within a fully specified equilibrium model. We make no such equilibrium assumption. Consistent with the atheoretical testing tradition noted above, we treat FIRE solely as the null hypothesis of an efficiency test and take no stand on the structural data-generating process or on whether any agent’s expectations are model-consistent.

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. Unlike predictable errors, big ex post 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 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 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 et al. (2020), Bordalo et al. (2019), Nagel and Xu (2022), Afrouzi, Kwon, Landier, Ma and Thesmar (2023), Bordalo, Gennaioli, La Porta and Shleifer (2024), De La O and Meyers (2021, 2023) Hillenbrand and McCarthy (2021).) Other researchers have argued that investors may be rationally or behaviorally inattentive or “boundedly rational” because they are incapable of attending to or processing all available

information at a given moment, leading them to underreact to some forms of information (e.g., Mankiw and Reis (2002), Woodford (2002), Sims (2003), Gabaix (2019), Kohlhas and Walther (2021)). A common approach in this literature investigates whether measured subjective beliefs depart from FIRE by regressing survey forecast errors on lagged forecast revisions. An alternative approach, used by Chen, Hansen and Hansen (2020), derives bounds on investor expectations using model-implied moment restrictions and constraints on statistical divergence from rational expectations. We extend this literature with 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. At the same time, placing distortions in a real-world context requires a practical measure of objective beliefs rather than a model-based benchmark. Model based benchmarks provide idealized parametric counterfactuals, but typically assume the data-generating process (DGP) is specified without error and that agents in the model perfectly observe this process and can rapidly update their understanding of its parameters in the wake of structural change. In reality, the true DGP is inherently unknown and potentially unknowable with structural change inevitable. Our machine methodology allows us to agnostically estimate the benchmark directly from the data while seeking to learn about structural change, ensuring robust empirical description without rigid functional form assumptions. 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, Décaire and Guenzel (2024), who build a mental model of sell-side analysts’ forecasts that is informed by a large language model analysis of their written reports. Their historical analysis differs from our real-time analysis and does not share our goal of forming a measure of objective expectation formation. Despite the differing methodologies, our findings exhibit key similarities, including 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 predictable mistakes in subjective beliefs. A related adaptive machine learning approach has recently been applied to forecasting inflation from microdata rather than financial returns (Chen, Gao, Hazell, Lei and Lian (2026)).

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 subsamples and is written generically as:

$$y_{j,t+h} = G(\mathcal{X}^t, \beta_{j,h,t}) + \epsilon_{jt+h}. \quad (1)$$

where $\mathcal{X}^t$ denotes the history of data inputs from time $1, \dots, t$ 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 $\beta_{j,h,t}$. Our main approach uses the Long Short-Term Memory (LSTM) deep sequence recurrent neural network for $G(\mathcal{X}^t, \beta_{j,h,t})$.⁵ 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 $\beta_{j,h,t}$ is used to denote one in a sequence of time-invariant parameter estimates obtained from rolling subsamples, rather than estimates that vary over time within a sample. Estimation of (1) delivers a time t machine “belief” about $y_{j,t+h}$ given by the estimated value of $G(\mathcal{X}^t, \beta_{j,h,t})$. We denote this time t objective belief as $\mathbb{E}_t[y_{j,t+h}]$.

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 subsample windows: a “training” subsample 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

⁵The LSTM estimator with N hidden layers $h_t^N \in \mathbb{R}^{D_{h^N}}$ takes the form

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

See the Online Appendix for the detailed specification.

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 each 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 pseudo 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, window lengths, and network architecture structure. 4) Out-of-sample prediction. In this step a true 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 $\{\mathbb{E}_t[y_{j,t+h}], \mathbb{E}_{t+1}[y_{j,t+1+h}], \dots, \mathbb{E}_T[y_{j,T}]\}$.

Several points about the algorithm bear emphasizing. First, the machine forecast is made without referencing the time- t survey forecast, but the estimate of prestakes is survey-specific. In particular, the algorithm checks period-by-period whether the specification choice made by the machine would have judged a given survey forecast to be predictably in error based on a comparison of pseudo forecast errors over the validation sample that ends one-period before the machine issues its own forecast. Our period-by-period measure of prestakes in subjective beliefs measured as $bias_{s,t}[y_{t+h}] \equiv \mathbb{F}_{s,t}[y_{t+h}] - \mathbb{E}_t[y_{t+h}]$ if the ex ante check detects predictable mistakes, and zero otherwise. It is worth emphasizing that 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, they are not included in $bias_{s,t}$. In particular, bias in expectations is measured relative to the ex ante machine forecast, not relative to an ex post outcome. It follows 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.

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 $T_E \in \{\underline{T}_E, \dots, \bar{T}_E\}$ $T_V \in \{\underline{T}_V, \dots, \bar{T}_V\}$ 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.

Third, 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 $t = (\dot{T} + h), \dots, T$. The accumulation of true out-of-sample forecast errors serves as an external validation step that exists outside of the optimization loop. Although we need to reserve a minimum number of past observations for training and hyperparameter tuning, 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 balance these considerations by requiring that our testing sample cover an extended time-frame from 2004:Q1 to 2023:Q4, reserving the pre-2004 sample for training and validation. Because return forecasts are made four quarters ahead, after accounting for this horizon the final four-quarter period of the testing sample runs from 2023:Q1 to the end of 2023:Q4.

Machine Input Data 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 our data and sources 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.⁶ 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 93 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,⁷ and daily text-based factors (topic intensities) estimated by Latent Dirichlet Allocation (LDA) analysis from around one million articles published in the *Wall Street Journal* between January 1984 to June 2024.⁸

We find it useful to separate the large number of inputs as falling into one of three coarse categories: “macro fundamentals,” “non-fundamentals,” and “media-sentiment.” Macro fundamentals include data or factors that capture variation in employment, hours, output, corporate earnings, dividends, 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 non-fundamentals because their longer duration means they are influenced by both subjec-

⁶It 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.

⁷That fed funds futures jump around FOMC announcements was publicly known at least five years before our testing sample begins, following the circulation of Kuttner (2000).

⁸Following Bybee et al. (2021) the LDA factor estimation is based on a randomly selected subsample of 200,000 articles over the same period.

tive 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 H.1 in the Online Appendix.

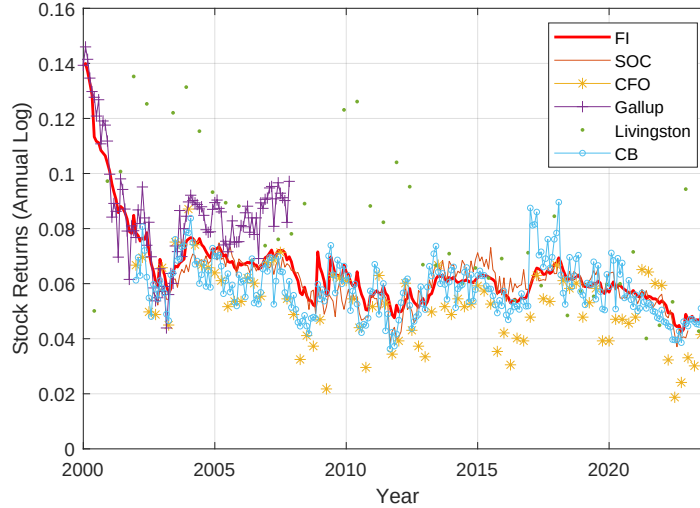
Survey Data For survey expectations of returns, we use mean forecasts from Gallup/UBS, the Michigan Survey of Consumers (SOC), the Conference Board (CB), the CFO Survey from the Richmond Federal Reserve Bank, the Livingston survey, and the Value Line survey.⁹ Surveys vary according to whether they ask 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.¹⁰ 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 D 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 noisy surveys

⁹We thank David Thesmar for generously providing their Value Line expected returns data. For details on data construction, see Thesmar and Verner (2025).

¹⁰The SOC asks about the probability of a rise in the market and the CB asks for a categorical belief on whether the market will rise, stay the same, or fall. The Appendix explains how we convert these into a quantitative point forecast.

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

Figure 1: Filtered Investor Expected Return Series (FI)



Notes: Figure plots one-year-ahead filtered investor return expectations (FI) extracted using a state-space model that combines multiple investor forecasts (Gallup/UBS, CFO, Livingston, SOC, and CB) 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).

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 D for details). We measure earnings expectations with the median IBES equity analyst forecasts of earnings growth over the next four quarters, i.e., one-year-ahead earnings growth on the S&P 500. We also consider the IBES median long-term growth forecast (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 three possibilities for the purpose of evaluating forecast errors.¹¹ First,

¹¹For 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

we interpret LTG as measuring annual five-year forward growth expectations, i.e., annual earnings growth from four to five years ahead. This interpretation is motivated by Bianchi et al. (2026), who compared the mean squared forecast errors using LTG to predict different 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. Second, 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. This interpretation follows the implicit assumption in Nagel and Xu (2022). Third, we follow Bordalo et al. (2024) and interpret LTG as measuring the five-year growth rate, i.e., from 0 to five years ahead.

Realized Outcome Data Our realized outcome data represent the forecast targets of the survey expectations data. For realized stock market returns, we use the annual Center for Research in Securities Prices value-weighted return series (CRSP-VW, includes dividends). For realized earnings, we compute a bottom up measure of S&P 500 “street earnings” from IBES. Street earnings differ from GAAP earnings by excluding discontinued operations, extraordinary charges, and other non-operating items and has been emphasized as the appropriate forecast target for IBES equity analyst earnings forecasts.¹²

3 Results

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

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.

¹²According 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.

3.1 Preliminary Analysis

Table 1 summarizes the forecasting performance of the unadjusted machine algorithm 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., one-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.

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 four-quarter-ahead surveys of the stock market, plus our FI series: the Gallup/UBS poll of individual investors, 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). We separately compare the machine to the Value Line (VL) survey of equity analysts, whose return forecasts are made at a much longer, sixteen-quarter horizon. 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 of 8, 12, and 16 quarters ahead. Panel (a) shows that the machine algorithm produces return forecasts that are substantially more accurate than all of these four-quarter-ahead benchmarks. The ratio of the machine-MSE to survey-MSE is 0.68, 0.66, 0.79, 0.64, 0.75, and 0.67 when compared to the CB, SOC, Gallup/UBS, CFO, LIV, and FI surveys, respectively. This corresponds to out-of-sample R^2 statistics of 0.32, 0.34, 0.21, 0.36, 0.25, and 0.33, respectively, implying that the machine forecasts are 32%, 34%, 21%, 36%, 25%, and 33% more accurate than the respective investor forecasts. At the

sixteen-quarter horizon, the ratio of the machine-MSE to VL-MSE is 0.73, corresponding to an out-of-sample R^2 of 0.27, or 27% more accurate than the VL survey. The machine is also substantially more accurate than the historical mean, with the ratio of the machine MSE to historical mean MSE equal to 0.68, 0.70, 0.72 and 0.83 for one-year, two-year, three-year, and four-year return forecasts. The advantage does not extend to every horizon: at five years the ratio is 1.19, so the machine no longer improves on the historical mean.

Table 1: Forecasting Performance: Machine vs. Benchmarks

Mean-square-errors of machine forecasts (MSE_E) versus survey/benchmark (MSE_F)							
(a) Stock Returns							
Horizon h (quarters)	4	4	4	4	4	4	16
Survey/benchmark	CB	SOC	Gallup/UBS	CFO	Livingston	FI	VL
MSE_E/MSE_F	0.681	0.663	0.788	0.638	0.745	0.666	0.729
OOS R^2	0.319	0.337	0.212	0.362	0.255	0.334	0.271
w^*	1.000	1.000	1.000	1.000	0.999	1.000	0.653
MSE_F/MSE_H (Mean)	0.998	1.027	0.999	1.066	0.968	1.022	1.144
Horizon h (quarters)	4	8	12	16			
Survey/benchmark	Mean	Mean	Mean	Mean			
MSE_E/MSE_F	0.680	0.697	0.719	0.834			
OOS R^2	0.320	0.303	0.281	0.166			
w^*	1.000	1.000	0.927	0.664			
(b) S&P 500 Earnings Growth							
Horizon h (quarters)	4	LTG 4-5	LTG 1-10	LTG 0-5			
Survey/benchmark	IBES	IBES	IBES	IBES			
MSE_E/MSE_F	0.386	0.824	0.164	0.382			
OOS R^2	0.614	0.176	0.836	0.618			
w^*	1.000	1.000	1.000	0.999			
MSE_F/MSE_H (Mean)	1.599	1.143	3.972	2.463			

Notes: MSE_E , MSE_F , and MSE_H are the mean-squared-forecast-errors of machine, benchmark, and historical mean forecasts, respectively. OOS R^2 is defined as $1 - MSE_E/MSE_F$. The weight on the machine forecast 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 CRSP stock returns and against realized log CRSP stock returns. Filtered Investor (FI) expectations series aggregates forecasts from Gallup/UBS, CFO, Livingston, SOC, and CB surveys into a latent one-year-ahead expected return using a state-space model. Value Line (VL) expected returns are the annualized nominal expected total return from Thesmar and Verner (2025). The testing subsample is 2005:Q1 to 2023:Q4 for stock returns when the benchmark is CB, SOC, CFO, FI, and VL, 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 subsample is 2005:Q1 to 2023:Q4, with h -quarter forecasts constructed from the middle month of each quarter. LTG 4-5, LTG 1-10, and LTG 0-5 denote Long-Term Growth (LTG) forecasts that are interpreted as earnings growth for 4-to-5, 1-to-10, and 0-to-5 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.

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 as

described above. 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 0.39 for one-year-ahead earnings growth, implying the machine is 61% more accurate than the median analyst for these predictions. The machine is about 18% 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 statistical uncertainty. 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 w^* would be 1

if the machine were always better than the survey, 0.5 if the machine and survey forecasts were similarly good, so that one would want to place equal weight on the two, and 0 if the machine were worse than the survey on average. In almost 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. The VL survey is the exception, with w^* equal to 0.653: the machine still does better, but by a smaller margin, consistent with VL's longer, 16-quarter forecast horizon, where mean reversion in returns is likely to dominate and leave less room for a forecasting advantage.

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 1-year survey forecasts are negatively correlated with the subsequent realized returns over the period being forecast. The exceptions are the Livingston and Conference Board surveys, whose forecasts have correlations of 0.19 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 0.58 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 more negative than the correlation between the surveys themselves and subsequent returns, with the exception of the Gallup/UBS survey. 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 -0.34 for the Livingston survey to a low of -0.58 for the Gallup/UBS survey. This shows that once we isolate times of predictable *over*-optimism (as distinct from optimism per se) we find much larger inaccuracies. This suggests that prestakes drive the negative correlation between survey return measures and subsequent returns.

Table 2: Correlations of stock return forecasts and bias with realized returns

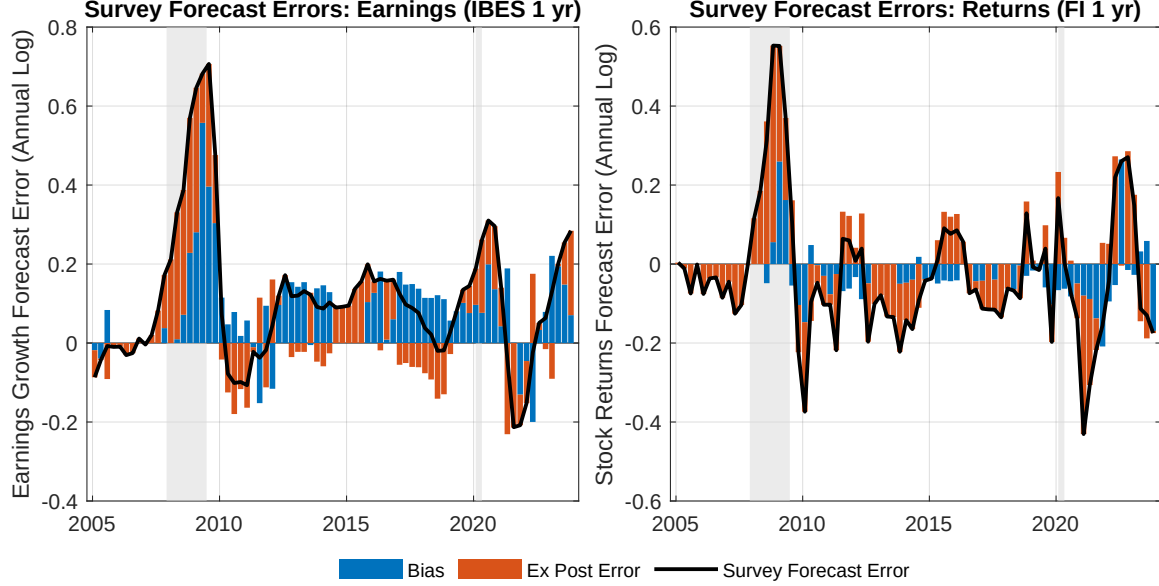
(A) Raw return forecasts vs. realized returns	
Forecast	Correlation
Filtered Investor	-0.173
CFO	-0.242
Livingston	0.191
Gallup/UBS	-0.641
SOC	-0.183
CB	0.128
Value Line	0.213
Machine	0.578
(B) Raw return bias vs. realized raw returns	
Forecast	Correlation
Filtered Investor	-0.544
CFO	-0.550
Livingston	-0.335
Gallup/UBS	-0.579
SOC	-0.543
CB	-0.513
Value Line	-0.144

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, where bias is defined as the survey forecast minus the machine forecast when prestakes are found. The sample covers 2005:Q1-2023:Q4. Forecasts include investor expectations (filtered investor, CFO, Livingston, Gallup/UBS, SOC, Conference Board, Value Line), and the machine learning forecast. Value Line is a sixteen-quarter (three-to-five-year) forecast and is correlated with the corresponding sixteen-quarter realized return; the other forecasts are four-quarter ahead.

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 decomposition of survey forecast error into the prestake component and an ex post error component. Specifically, in periods where prestakes are found, $\mathbb{F}_{s,t}[y_{t+h}] - y_{t+h} = (\mathbb{F}_{s,t}[y_{t+h}] - \mathbb{E}_t[y_{t+h}]) + (\mathbb{E}_t[y_{t+h}] - y_{t+h})$, where the first term is the period-specific prestake $\text{bias}_{s,t}[y_{t+h}] \equiv \mathbb{F}_{s,t}[y_{t+h}] - \mathbb{E}_t[y_{t+h}]$, and second term is the ex post error $\mathbb{E}_t[y_{t+h}] - y_{t+h}$. If no prestakes are found, the first term is zero and the survey forecast error is entirely comprised of ex post error. The results for IBES expectations of earnings growth are shown on the left, while those for the F1 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 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. This string of ex post forecast errors, as opposed to prestakes, implies 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.

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). NBER recessions are shaded in gray. The sample spans 2005:Q1-2023:Q4. The x-axis represents the date at which the forecast target is realized. Forecasts (machine and survey) are plotted alongside the actual realized values over the corresponding period being forecast, so that a vertical slice shows the forecast error.

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:

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

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 $t - 1$. 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, $\hat{\beta}_t$, 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 $\mathcal{X}_t$. Here, $\nabla G(\cdot)$ denotes the integrated gradient of the LSTM function with respect to $\mathcal{X}$, evaluated at $\mathcal{X}_t$ and $\mathcal{X}^{t-1}$. 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 G for details.

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 LSTM estimator at time t , using the history of input data through $t - 1$. To mitigate over-fitting, we estimate recursive out-of-sample forecasts as we do for the target variable forecasts. We then decompose the time t survey forecast into the same components as the machine forecast, plus a residual:

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

The residual captures the error in the fitted model and represents input features intangible to the machine, such as private information or forecaster judgment.¹³

Figure 3 gives a visual impression of the forecasts and their decompositions into broad

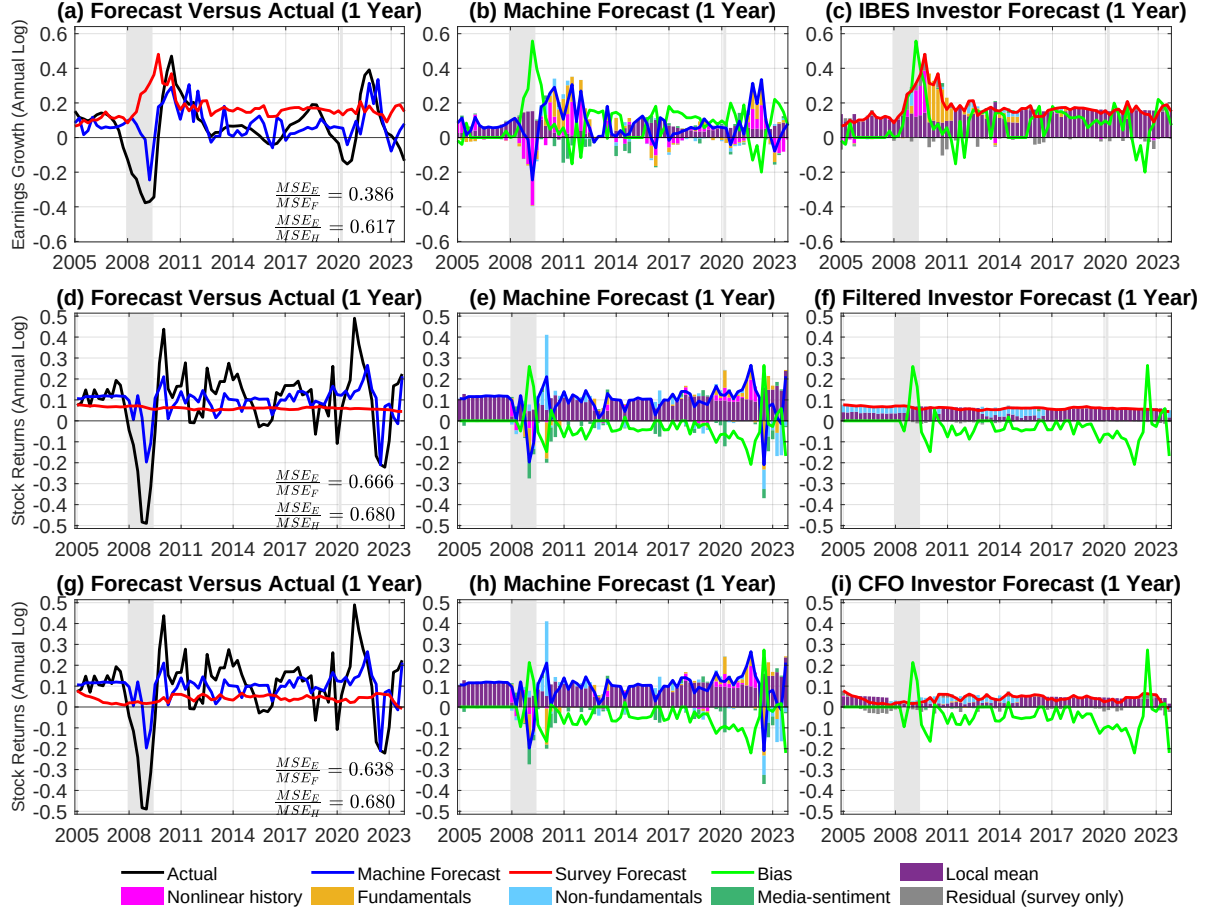
¹³In principle the error also contains the approximation error for the decomposition, but the magnitude of approximation error is negligible.

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%.¹⁴

Figure 3 shows that both surveys of earnings growth and returns—but especially subjective return expectations—are extremely stable relative to the volatile realized series that represent what respondents are targeting. This finding 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, and other surveys are similarly stable. IBES earnings growth expectations are more volatile, but almost never predict negative earnings growth. The third column shows that survey forecasts for both earnings growth and returns closely track the slowly moving local mean much of the time. These local mean components largely reflect extrapolation and are highly correlated with moving averages of the target variables. The exceptions to this for IBES earnings growth forecasts are largely confined to the period 2010-2011, where fundamentals play a large role, and for the survey return forecasts to the periods 2005-2008 and 2013-2016, where non-fundamentals (i.e., financial market indicators) play a large role. By contrast, the machine forecasts deviate more often from its own local mean at various points over the full sample, with the nonlinear history, macro fundamentals, and media sentiment playing larger roles, especially during periods of important economic change. Figure 3 also shows a large variation in the survey $bias_t$ terms over the full testing sample. While it is known that IBES expectations can be biased upward on average, our interest is in the time-variation in these distortions, especially during periods of unusual

¹⁴Since 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.

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



Notes: Figure plots the machine and survey forecast and their decomposition into local means, nonlinear histories, and broad categories of current information variables. For the investor forecast, an additional residual term captures the portion of the forecast not explained by the LSTM model out of sample. MSE_E , MSE_F , and MSE_{FI} 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 returns is either the Filtered Investor (FI) or median CFO forecast. An additional residual term for the investor forecasts captures the portion of the forecast not explained by the LSTM model out of sample. Stock returns are expressed as log annual returns; as in the forecasting-performance tables, this decomposition is of the raw-return forecast, so no one-year T-bill rate is subtracted. NBER recessions are shaded in gray. The testing sample spans 2005:Q1-2023:Q4. The x-axis represents the date at which the forecast target is realized. Forecasts (machine and survey) are aligned with the realized values over the corresponding period being forecast, so that a vertical slice shows the forecast error.

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.

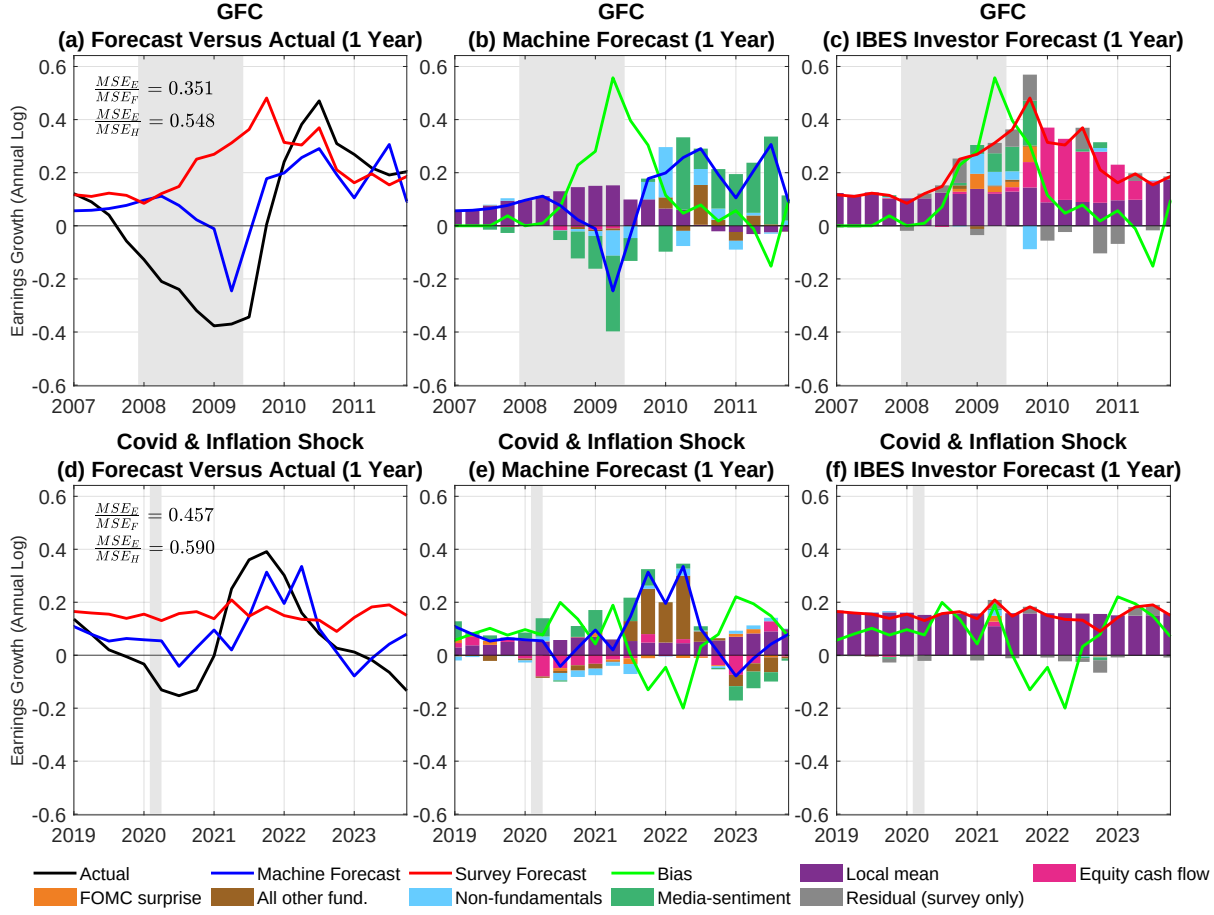
Figures 4 and 5 refine the decomposition used in Figure 3 in three respects. First, whereas Figure 3 reports the nonlinear-history component $h_t(\cdot)$ as a single aggregate bar, in Figures 4 and 5 we decompose $h_t(\cdot)$ itself by integrated gradients with respect to the input history, summing each input’s attribution across its lags to obtain one contribution per feature.¹⁵ Second, we merge this nonlinear-history attribution with the integrated-gradient decomposition of the current-information term $\nabla G(\cdot)' \mathcal{X}_t$, so that each feature’s total contribution—through both current information and the history—is assigned to a common set of economic themes; the nonlinear history therefore no longer appears as a separate bar but is distributed across those themes. Third, we disaggregate the coarse “fundamentals” theme into three economically interpretable sub-themes—equity cash-flow indicators, FOMC (monetary-policy) surprises, and all other fundamentals—while the remaining themes (non-fundamentals and media sentiment) are defined as in Figure 3.¹⁶

Several points about these figures bear noting. First, even in these unusual sub-periods, 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 suffer from a chronic over-optimism that is 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

¹⁵Contributions are reported at the level of economic themes rather than individual features. Several of our inputs are closely related to one another, so theme-level totals are measured more precisely than the allocation among features within a theme.

¹⁶The equity cash-flow theme collects the inputs most directly tied to expected corporate cash flows: lagged realized earnings growth, lagged 1-year IBES earnings growth forecast, a cyclical (i.e., linearly detrended) component of the survey/analyst earnings-level forecast, a moving average of earnings growth, the IBES consensus long-term earnings-growth expectation (LTG), and the earnings-GDP ratio. FOMC surprises are built from high-frequency changes in fed funds and Eurodollar futures around FOMC announcements.

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

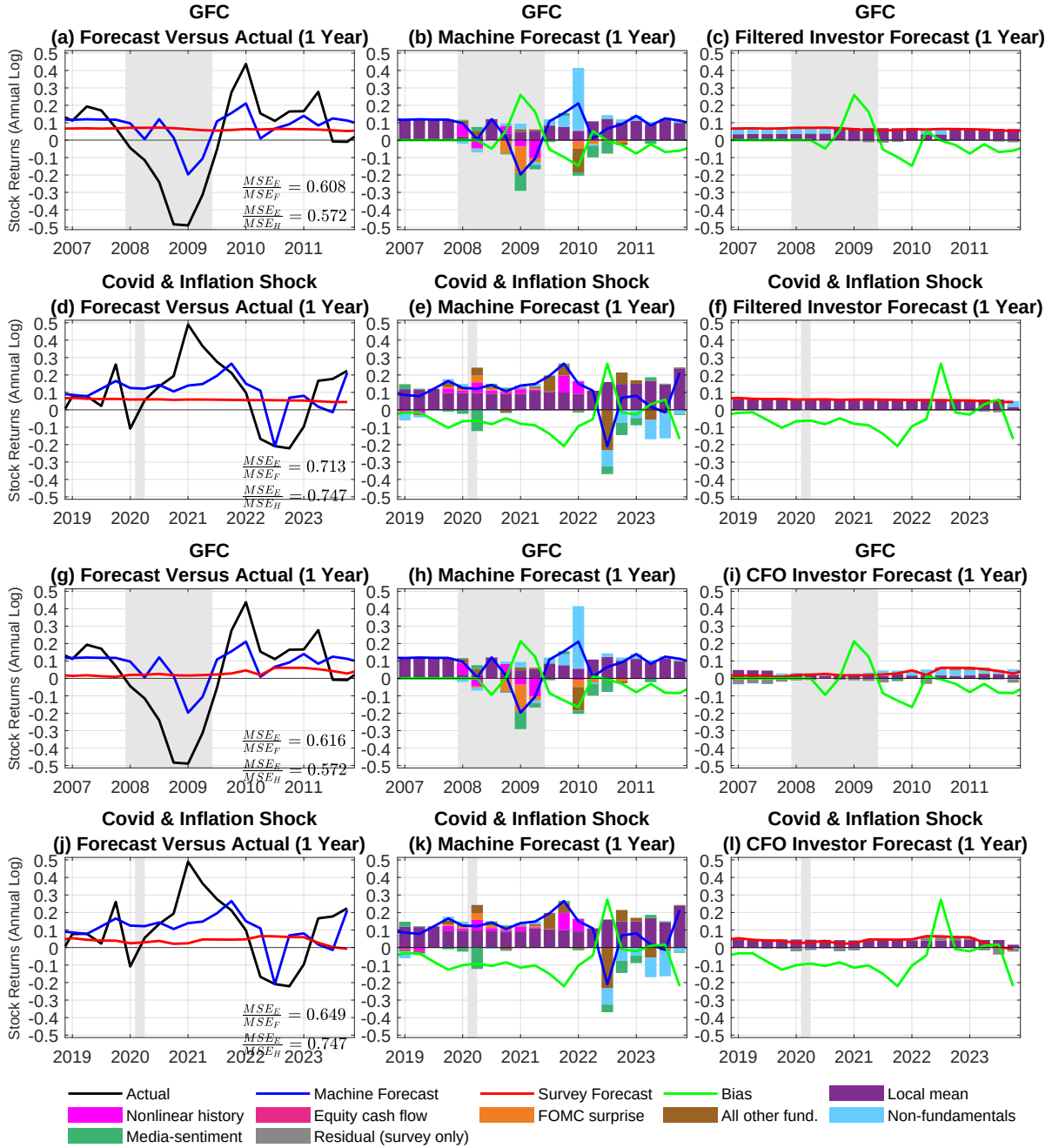


Notes: See Figure 3. The figure plots crisis-window decompositions of the machine and IBES investor forecasts by reporting a lag-aggregated integrated gradient decomposition of the non-linear histories, merged into common themes with the current information components. The subsample plots span 2007:Q1-2011:Q4 (GFC) and 2019:Q1-2023:Q4 (Covid-19 & Inflation Shock).

measure of objective beliefs, which is not appreciably over-optimistic on average, precisely because it places greater weight on the nonlinear history, macro fundamentals, and media sentiment variables especially in crisis episodes in order to identify turning points. Second, Figure 4 shows that, during the GFC, the IBES analysts' mounting over-optimism (panel (c)) is driven largely by rising media-sentiment attributions. Yet the very same topic intensities first push the machine toward greater pessimism before turning it optimistic, tracing the V-shaped forecast (panel (b)) that tracks the collapse and rebound in realized earnings far more accurately. The machine thus reads the same media signals not as grounds for optimism but as a marker of conditions that breed over-optimism in earnings analysts. 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 previous times of unusual volatility. The same split shows up in the fundamentals: aside from equity cash-flow indicators, broader macro fundamentals scarcely register in the survey forecast, even as they drive much of the machine's through the Covid and post-Covid bear market (panel (e)) and play a smaller part in the 2010–11 recovery. Third, both 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 evidence that an economic crash is underway. 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. For example, 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

Figure 5: Analysis of Stock Return Expectations: Crisis Episodes



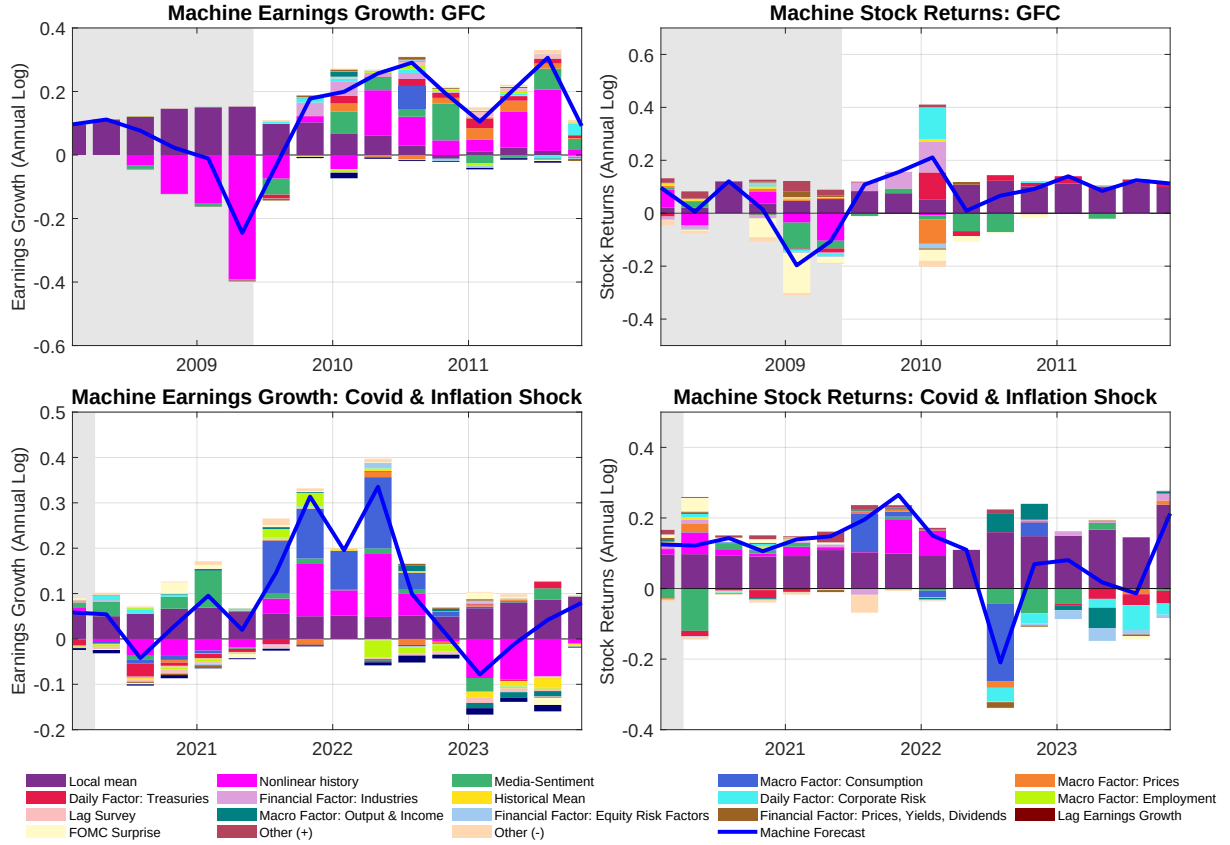
Notes: See Figure 3. The figure plots crisis-window decompositions of the machine and investor forecasts by reporting a lag-aggregated integrated gradient decomposition of the non-linear histories, merged into common themes with the current information components. The subsample plots span 2007:Q1-2011:Q4 (GFC) and 2019:Q1-2023:Q4 (Covid-19 & Inflation Shock).

these episodes, Figure 6 zooms in on the most important current information features in the machine specification across all input variables and themes. For return forecasts, the nadir of the GFC objective belief forecast occurs in 2009:Q1, and is driven mostly by media-sentiment and the FOMC surprise 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 including the 2022 bear market, macro factors related to consumer spending played important roles in predicting the rise and subsequent decline in both objective earnings growth and return forecasts.

Taken together, Figures 3 and 4 show that the GFC episode is the single most quantitatively important instance of systematic bias in analyst earnings expectations over the course of our sample. Not only do survey responses fail to predict any meaningful part of the plunge in earnings growth during the crisis, survey expectations veer sharply *positive* even as the machine forecast turns sharply negative and realized earnings growth would continue its downward trajectory for over a year. To study the reasons for the predictable mistakes in this episode, we zero-in on the drivers of the bias in IBES expectations during the GFC and its aftermath.

To start, a comparison of panels (b) and (c) of Figure 3 shows that the same features often drive the machine forecast and IBES survey forecast, but in opposite directions. This is true of topic intensities from media reports, as previously noted, but also true for non-fundamental financial market indicators. Though not shown in the figure, further unpacking the non-fundamentals category indicates that spiking corporate-credit/default risk and collapsing commodity prices lead the machine to mark earnings growth sharply down in late 2008 and early 2009, while the survey reads the very same series as grounds to mark earnings up. During the recovery in 2010–11 realized earnings growth stayed strong and kept rising, and the machine—reading improving credit conditions and the topic intensities that signal expansion—likewise kept its growth forecast elevated and rising. Yet the IBES growth fore-

Figure 6: Detailed Machine Forecast Decompositions



Notes: The figure plots the machine forecast and its decomposition into core themes. Taken together, the components represented with colored bars account for 100% of the machine forecast in each period. Stock returns are expressed as log annual returns and earnings growth 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). The x-axis represents the date at which the forecast target is realized.

cast instead turned down in this period, which the decomposition pins the survey’s decline almost entirely on its equity cash-flow signal—specifically the IBES LTG expectation, which analysts marked steadily lower as the recovery matured. Analysts appear to have been selectively attentive to the cyclical component in earnings growth, to the exclusion of almost all other macroeconomic indicators. The end result is that the IBES forecast predicted the rebound in earnings growth roughly a year too early, while missing its collapse entirely.

This evidence, especially when combined with that in panel (c) of Figure 3 showing that most everywhere else in our sample IBES forecasts are largely explained by a local mean in past earnings growth, underscores the extent to which IBES analysts are largely selectively focused on earnings-related information, while being sub-optimally attentive to broader macroeconomic and financial market data. These results for aggregate earnings expectations echo the firm-level findings of Bastianello et al. (2024) who study IBES reports and conclude that analysts overreact to firm-specific topics while underreacting to macro-related topics.

3.3 Implications for Over-/Underreaction

Tests of FIRE conventionally regress the survey forecast error, $y_{t+h} - \mathbb{F}_t[y_{t+h}]$, on the forecaster’s own revision, and read a positive slope as underreaction and a negative slope as overreaction. We ask what our estimate of bias—the object this paper is built to measure—implies for that same margin. Because $bias_t[y_{t+h}] \equiv \mathbb{F}_t[y_{t+h}] - \mathbb{E}_t[y_{t+h}]$ runs opposite in sign to a forecast error, we place its negative on the left-hand side,

$$\text{NEG bias}_{IBES,t}[y_{t+h}] \equiv \mathbb{E}_t[y_{t+h}] - \mathbb{F}_{IBES,t}[y_{t+h}],$$

when a prestake has been found. This quantity is the predictable component of the survey forecast error, so it inherits the literature’s orientation along with its familiar sign convention: a positive coefficient on a revision signals underreaction, a negative coefficient overreaction.

Let $\mathbb{X}_t$ generically denote a forecast, e.g., either $\mathbb{E}_t$ or $\mathbb{F}_t$. Table 3 reports estimates of

$$\text{NEG bias}_{IBES,t}[y_{t+h}] = \alpha + \beta (\mathbb{X}_t - \mathbb{X}_{t-4}) [y_{t+h}] + u_t$$

for three forecast objects, varying the four-quarter forecast revision $\mathbb{X}_t - \mathbb{X}_{t-4}$. Column (1) uses the machine’s revision $\mathbb{E}_t - \mathbb{E}_{t-4}$; column (2) uses the analysts’ own $\mathbb{F}_{IBES,t} - \mathbb{F}_{IBES,t-4}$. Every coefficient in column (1) is positive and every coefficient in column (2) is negative, with 90% credible sets bounded away from zero. When objective beliefs are revised up, subjective beliefs move up by less and the wedge between them widens, implying underreaction to objective news. By contrast, when analysts revise up, their forecasts move further than objective beliefs do and the wedge narrows, implying overreaction to perceived news.¹⁷

Bordalo et al. (2020) propose a mechanism for this pair of signs, reconciling them within a diagnostic expectations version of a dispersed information learning model: forecasters overreact to their own noisy signals, which turns the individual coefficient negative, and react not at all to the information received by others, which turns the consensus coefficient positive. Our results provide a natural test of that mechanism. Their comparison runs from the individual analyst to the analyst consensus, ours from the analyst consensus to the machine, which aggregates approximately all signals relevant to the outcome; it is effectively the same mechanism applied at two different levels of information aggregation. Since IBES analysts attend to earnings-related information and largely ignore the macroeconomic, financial market, and media signals the machine uses, column (2) records overreaction to the signals they attend to and column (1) their failure to react to the rest, as the mechanism predicts.

¹⁷These findings are robust to adding the four-quarter lag of realized earnings growth as a second regressor. This nets out persistence in earnings growth, so that the coefficient on the revision reflects reaction to news rather than mechanical predictability. All six coefficients keep their signs and credible sets that exclude zero. Five are essentially unchanged; in row (c) of column (1) the machine-revision coefficient rises from 0.059 to 0.093. The pattern documented here is therefore not a mechanical consequence of persistent earnings growth.

Table 3: Implications of Prestakes for Over- and Underreaction

NEG bias $_{IBES,t}[y_{t+h}] = \alpha + \beta \cdot (\mathbb{X}_t - \mathbb{X}_{t-4}) [y_{t+h}] + u_t$		
	(1) Machine revision $\mathbb{X}_t = \mathbb{E}_t$	(2) Survey revision $\mathbb{X}_t = \mathbb{F}_{IBES,t}$
(a) 0-to-5 yr LTG		
$\hat{\beta}$	+0.010	-0.011
90% Credible Set	[+0.005, +0.014]	[-0.016, -0.007]
Adj. R^2 (%)	16.2	18.5
N	72	76
(b) 4-to-5 yr LTG		
$\hat{\beta}$	+0.021	-0.011
90% Credible Set	[+0.014, +0.028]	[-0.019, -0.003]
Adj. R^2 (%)	24.3	5.7
N	72	76
(c) 1-yr earnings		
$\hat{\beta}$	+0.059	-0.071
90% Credible Set	[+0.040, +0.079]	[-0.088, -0.054]
Adj. R^2 (%)	25.7	38.2
N	72	76

Notes: The table reports estimates of the regression in the header, where NEG bias is the machine forecast minus the survey forecast, $\text{NEG bias}_{IBES,t}[y_{t+h}] \equiv \mathbb{E}_t[y_{t+h}] - \mathbb{F}_{IBES,t}[y_{t+h}]$, when a prestake has been found. Column (1) uses the machine forecast revision, and column (2) uses the survey forecast revision. Row (a) uses the 0-to-5 year cumulative LTG forecast, row (b) the 4-to-5 year forward LTG forecast, and row (c) the one-year-ahead earnings growth forecast. Revisions are one-year (four-quarter) changes. Regressors are standardized, so coefficients are expressed per one-standard-deviation revision. 90% Bayesian credible sets are the closed-form flat-prior posterior intervals, the exact limit of the 100,000 posterior t -draws used for the ratio and alpha tables. The sample spans forecast target dates 2005:Q1-2023:Q4. The machine forecast begins with the testing sample, so its revision is unavailable for the first four target dates, leaving 72 observations in column (1) and 76 in column (2).

3.4 Subjective vs. Objective Risk-Return Trade-off

Any finding of meaningful predictable mistakes in subjective beliefs raises the possibility that real-world investors have a distorted view of the objective trade-off between risk and reward. To consider this possibility, we investigate how the risk-return trade-off 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 the CFO survey and provides a quarterly measure 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 D 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 4 reports regressions of subjective (objective) expected excess returns on subjective (objective) measures of risk. Since the risk measures are on different scales, we normalize by dividing each risk measure by its own standard deviation, so that each coefficient reflects the response to a one-standard-deviation change in that 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 4 shows that we find a positive risk-return trade-off in both subjective and objective beliefs. However, we find a large difference in magnitudes. Objective risk premia move between 1.9 and 8.1 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, the posterior probability that the subjective risk coefficient is less than the objective risk coefficient exceeds 82% in all cases.¹⁸ 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 4 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 trade-off with the objective risk-return trade-off. 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 95% 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.

3.5 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 are based on information that we verify would have been available to survey respondents but are systematically more accurate, 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

¹⁸This can happen because some draws from the posterior will include small values for the denominator coefficient.

Table 4: Subjective vs. Objective Risk-Return Trade-off

	Survey: $\mathbb{F}_t^s[r_{t,t+h}^e] = \beta_{\mathbb{F},0} + \tilde{\beta}_{\mathbb{F},1} \tilde{\mathbb{V}}_t[r_{t,t+h}] + \epsilon_{\mathbb{F},t}$	
	Machine: $\mathbb{E}_t[r_{t,t+h}^e] = \beta_{\mathbb{E},0} + \beta_{\mathbb{E},1} \mathbb{V}_t[r_{t,t+h}] + \epsilon_{\mathbb{E},t}$	
Survey s	Filtered Investor (FI)	Nagel and Xu (NX)
Horizon h (quarters)	4	4
(A) Perceived Risk $\tilde{\mathbb{V}}_t[r_{t,t+h}]$: Crash Confidence Index (Individuals)		
Objective Risk $\mathbb{V}_t[r_{t,t+h}]$: Financial Uncertainty Index		
$\tilde{\beta}_{\mathbb{F},1}/\beta_{\mathbb{E},1}$	0.273	0.317
90% Credible Set	$[-0.89, 1.42]$	$[-1.06, 1.67]$
$P(\tilde{\beta}_{\mathbb{F},1}/\beta_{\mathbb{E},1} < 1)$	0.924	0.907
(B) Perceived Risk $\tilde{\mathbb{V}}_t[r_{t,t+h}]$: Perceived Variance (CFO)		
Objective Risk $\mathbb{V}_t[r_{t,t+h}]$: Financial Uncertainty Index		
$\tilde{\beta}_{\mathbb{F},1}/\beta_{\mathbb{E},1}$	0.188	0.124
90% Credible Set	$[-0.63, 1.05]$	$[-0.69, 1.00]$
$P(\tilde{\beta}_{\mathbb{F},1}/\beta_{\mathbb{E},1} < 1)$	0.947	0.950
(C) Perceived Risk $\tilde{\mathbb{V}}_t[r_{t,t+h}]$: Crash Confidence Index (Institutions)		
Objective Risk $\mathbb{V}_t[r_{t,t+h}]$: Financial Uncertainty Index		
$\tilde{\beta}_{\mathbb{F},1}/\beta_{\mathbb{E},1}$	0.540	0.485
90% Credible Set	$[-1.93, 2.76]$	$[-1.71, 2.52]$
$P(\tilde{\beta}_{\mathbb{F},1}/\beta_{\mathbb{E},1} < 1)$	0.828	0.845
(D) Objective Risk $\mathbb{V}_t[r_{t,t+h}]$: Financial Uncertainty Index		
Survey: $\mathbb{F}_t^s[r_{t,t+h}^e] = \beta_{\mathbb{F},0} + \beta_{\mathbb{F},1} \mathbb{V}_t[r_{t,t+h}] + \epsilon_{\mathbb{F},t}$		
Machine: $\mathbb{E}_t[r_{t,t+h}^e] = \beta_{\mathbb{E},0} + \beta_{\mathbb{E},1} \mathbb{V}_t[r_{t,t+h}] + \epsilon_{\mathbb{E},t}$		
$\beta_{\mathbb{F},1}/\beta_{\mathbb{E},1}$	0.159	0.009
90% Credible Set	$[-0.56, 0.92]$	$[-0.78, 0.80]$
$P(\beta_{\mathbb{F},1}/\beta_{\mathbb{E},1} < 1)$	0.954	0.960

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 $\tilde{\beta}_{\mathbb{F},1}/\beta_{\mathbb{E},1}$ using 100,000 posterior draws, where the numerator and denominator slopes are drawn jointly from a bivariate Student-t whose covariance is the same-quarter cross-equation covariance of the two slopes over the quarters where both regressions are observed, and whose degrees of freedom (the overlap sample size minus two) enter through one shared chi-squared draw, following the seemingly-unrelated-regressions structure of Zellner (1962). The 90% credible set is defined by the 5th and 95th percentiles of the posterior distribution. $\mathbb{P}(\tilde{\beta}_{\mathbb{F},1}/\beta_{\mathbb{E},1} < 1)$ denotes the share of posterior draws where the ratio is less than one. $r_{t,t+h}^e$ denotes the h -quarter-ahead log return on the CRSP value-weighted index (including dividends) in excess of the T-bill. Subjective expectations $\mathbb{F}_t^s[r_{t,t+h}^e]$ are measured from surveys with $s = FI, NX$. Each risk measure is divided by its own standard deviation and the resulting coefficient is multiplied by 100, so it reflects the response (times 100) to a one-standard-deviation change in the corresponding risk measure. The sample spans 2005:Q1-2023:Q4.

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 market timing 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 market timing 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 the resulting one-year returns realized from 2005:Q1 to 2023:Q4. In order to exploit these predictions, the strategy makes its first portfolio choice in 2004:Q1 (the forecast of the one-year return realized in 2005:Q1 is formed in 2004:Q1) and uses 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 four quarters, we rebalance and the 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 where we start (2004:Q1, 2004:Q2, 2004:Q3 or 2004: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 $r_{s,t,t+4} = r_{t,t+4} \times \mathbf{1}_{\{\mathbb{P}_{s,t}[r_{t,t+4}] > r_{f,t}\}} + r_{f,t} \times \mathbf{1}_{\{0 < \mathbb{P}_{s,t}[r_{t,t+4}] \leq r_{f,t}\}} - (r_{t,t+4} - r_{f,t}) \times \mathbf{1}_{\{\mathbb{P}_{s,t}[r_{t,t+4}] \leq 0\}}$. 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 A.1 in the Online Appendix 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 earns risk-adjusted returns equal to 4.5 percent per annum relative to the CAPM and 4.9 percent per annum relative to the Fama-French 5-factor model. As a point of comparison, the perfect foresight alphas are 7.4 and 6.5 percent per annum for the two models, so the machine captures 60 percent of the perfect foresight alpha under the CAPM and 77 percent under the Fama-French 5-factor model. 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 almost never anticipate negative excess returns.

Table A.2 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 made on the basis of the machine forecast, this strategy would have earned alphas of 7.0 and 8.5 percent per annum relative to the CAPM and Fama-French 5-factor models, respectively. This may be compared to the infeasible perfect foresight alphas of 14.2 and 12.4 percent per

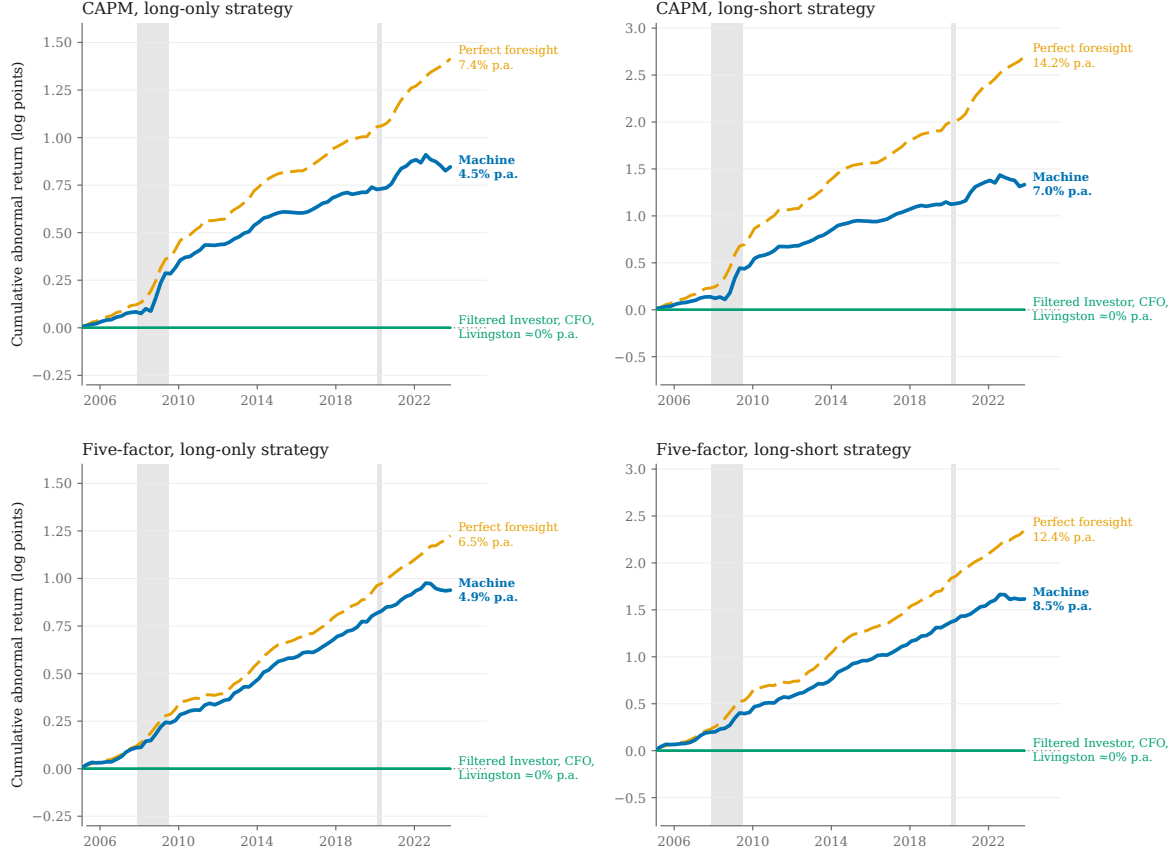
annum, so the machine captures 49 percent of the perfect foresight alpha under the CAPM and 69 percent under the Fama-French 5-factor model. When trades are made on the basis of the survey forecasts, alphas are again close to zero. This happens because the surveys almost never anticipate returns that fall below the T-bill rate, let alone losses. The strategies betting on objective beliefs all have market betas below unity, ranging from about 0.24 to 0.65.

An alpha estimate summarizes the average abnormal return over the testing sample but is silent about when that return was earned. We therefore accumulate the abnormal return through the sample, tracing in Figure 7 the path along which each strategy earns its alpha. That path separates an advantage confined to a single episode from one that decays as the sample proceeds and from one that accrues throughout.

The figure shows that abnormal returns accrue faster during crises than at other times. Taking the CAPM long-only strategy, 24 percent of its cumulative abnormal return arrives in the 2007–2009 recession, at an annualized rate of 13.6 percent against a full-sample average of 4.5 percent, and the line climbs again through the four quarters ending in mid-2022, which span that year’s bear market, before giving part of that back in 2023. The larger share, however, accrues outside these episodes. Fifty-three percent of the long-only total accrues between mid-2009 and the end of 2019, a stretch of more than ten years containing no recession at all, and it accrues there at an annualized rate of 4.3 percent, essentially the full-sample average. The long-short CAPM strategy behaves the same way, earning 6.7 percent per annum over that decade against a full-sample average of 7.0 percent.

The shape of this path reflects when and how the machine-based abnormal return is earned. The machine’s forecast of the four-quarter excess return falls to or below zero on seven of the 76 quarterly forecast dates in the testing period, moving the strategy to the one-year T-bill, and all seven are concentrated in the global financial crisis and the 2022 bear market. Five of these seven exits were followed by negative realized four-quarter excess returns, and they include the worst, second-worst, and third-worst four-quarter periods of the 76 in the testing sample, along with the exit taken in 2021:Q3 that avoided the 2022 bear

Figure 7: Cumulative Abnormal Returns from Trading Strategies



Notes: Each panel plots the cumulative abnormal return of a market timing strategy, $CAR_t = \frac{1}{J_s} \sum_{\tau \leq t} [r_{s,\tau,\tau+4} - r_{f,\tau} - \hat{\beta}'_{q(\tau)} F_\tau]$, where F_τ is the market excess return under the CAPM and the five Fama-French factors under the five-factor model, and $\hat{\beta}_q$ is estimated by ordinary least squares from $r_{s,t,t+4} - r_{f,t} = \alpha_q + \beta'_q F_t + \varepsilon_t$ on the observations formed in calendar quarter q , with $q(\tau)$ the formation quarter of the position realized at τ . Because forecasts are made four quarters ahead, the horizontal axis is dated by the target quarter, so a movement in a line at date τ reflects the abnormal return on a position taken four quarters earlier. The quantity plotted averages across the annually rebalanced portfolios formed in successive formation quarters, with $J_s = 4$ for the quarterly strategies and $J_s = 2$ for the semiannual Livingston strategy. Each observation is a four-quarter return, so the portfolios run in parallel and their undivided cumulative sum would add them all together; the leading factor of one over the number of portfolios converts it into the average that Tables A.1 and A.2 report. Each strategy is evaluated only in those quarters for which its own forecast exists. The restriction binds solely for Livingston, which is conducted semiannually and is therefore evaluated over the 38 quarters of the testing sample in which a Livingston forecast is available, as elsewhere in the paper. The remaining forecasts are available in all 76 quarters. The three survey lines are drawn as a single flat line at zero; their estimated cumulative abnormal returns never depart from zero by more than a tenth of a log point. Shaded bars denote NBER recessions. The sample spans 2005:Q1-2023:Q4.

market. The two later exits, in mid-2022, were followed by positive returns as the market recovered. In the remaining sixty-nine quarters, the strategy holds the market and captures its gains during this period. Two features of this pattern combine to produce the steady rise in the cumulative abnormal return, seen in the CAPM panels of the figure. The first is a market beta below unity, 0.65 in the long-only strategy and 0.45 in the long-short strategy, which reflects a small number of departures from the market. The second is the absence of false negative signals at other times, when the strategy stays exposed to the market's gains. The second feature is what delivers the steady rise in abnormal returns during the calm decade between the two crises, from mid-2009 to the end of 2019. Importantly, the absence of false signals does not reflect a forecast that merely tracks the historical mean. Across the 42 forecast origination dates from 2009:Q3 to 2019:Q4 the machine's forecast ranges from 1.5 to 16.6 percent and stays above the one-year rate throughout, coming within about one and a quarter percentage points of the threshold that would have moved the strategy out of the market. By contrast, a forecast based on the historical mean moves only between 8.8 and 9.5 percent.

The same figure makes the contrast with subjective beliefs immediate. In all four panels the three survey strategies lie on zero for the whole sample. They do so not because these strategies happen to break even, but because the survey forecasts almost never fall below the one-year rate. Over the 76 quarterly forecast dates of the testing period the FI forecast falls below the one-year rate in one quarter and the CFO forecast in two, all of them in 2022 when the one-year rate rose above four percent; the Livingston forecast never does so on any of the 38 semiannual dates at which it is observed. The Livingston strategy therefore holds the market in every quarter in which it is observed and its abnormal return is zero by construction, while the FI and CFO strategies move to the T-bill in those few quarters and earn five-factor alphas of -0.3 and 0.2 percent per annum, each with a 90 percent credible set containing zero. The infeasible perfect foresight line finishes above the other lines in every panel, and betting on objective beliefs captures between roughly a half and three-quarters of it across the two strategies and factor models.

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. The basic fact that spot prices are far more volatile than survey forecasts does not preclude this possibility. 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.

4 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 that survey responses exhibit a form of 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 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. DÉCAIRE, AND M. GUENZEL (2024): “Mental models and financial forecasts,” Jacobs Levy Equity Management Center for Quantitative Financial Research Paper.

- 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.
- (2026): “What Hundreds of Economic News Events Say About Belief Overreaction in the Stock Market,” *Review of Financial Studies*, forthcoming.
- 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. LA PORTA, AND A. SHLEIFER (2019): “Diagnostic expectations and stock returns,” *The Journal of Finance*, 74, 2839–2874.
- (2024): “Belief overreaction and stock market puzzles,” *Journal of Political Economy*, 132, 1450–1484.
- 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, C., C. GAO, J. HAZELL, L. LEI, AND C. LIAN (2026): “Forecasting Inflation with Microdata: An Adaptive Machine Learning Approach,” Working paper.
- CHEN, L., Z. DA, AND X. ZHAO (2013): “What drives stock price movements?” *The Review of Financial Studies*, 26, 841–876.
- CHEN, X., L. P. HANSEN, AND P. G. HANSEN (2020): “Robust identification of investor beliefs,” *Proceedings of the National Academy of Sciences*, 117, 33130–33140.

- COIBION, O., AND Y. GORODNICHENKO (2015): “Information rigidity and the expectations formation process: A simple framework and new facts,” *American Economic Review*, 105, 2644–78.
- 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. (1970): “Efficient Capital Markets: A Review of Theory and Empirical Work,” *Journal of Finance*, 25, 383–417.
- 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: Elsevier, 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.
- HAMILTON, J. D. (1994): *Time Series Analysis*, Princeton, NY: Princeton University Press.

- 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, 1177–1216.
- KOHLHAS, A. N., AND A. WALTHER (2021): “Asymmetric attention,” *American Economic Review*, 111, 2879–2925.
- KUTTNER, K. (2000): “Monetary policy surprises and interest rates: evidence from the Fed funds futures markets,” Technical report, Federal Reserve Bank of New York.
- LEVY, B. (2024): “Caution ahead: Numerical reasoning and look-ahead bias in AI models,” *Available at SSRN 5082861*.
- LUCAS, R. E. J. (1987): *Models of Business Cycles*, New York: Basil Blackwell.
- LUDVIGSON, S. C., S. MA, AND S. NG (2021): “Uncertainty and Business Cycles: Exogenous Impulse or Endogenous Response?” *American Economic Journal: Macroeconomics*, 13, 369–410.
- 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.
- MINCER, J., AND V. ZARNOWITZ (1969): “The Evaluation of Economic Forecasts,” in *Economic Forecasts and Expectations: Analysis of Forecasting Behavior and Performance* ed. by Mincer, J.: National Bureau of Economic Research, 3–46.

- MUTH, J. F. (1961): “Rational Expectations and the Theory of Price Movements,” *Econometrica*, 29, 315–335.
- 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,” *Journal of Financial Economics*, 150, 103713.
- NORDHAUS, W. D. (1987): “Forecasting Efficiency: Concepts and Applications,” *Review of Economics and Statistics*, 69, 667–674.
- SARKAR, S. K., AND K. VAFA (2024): “Lookahead bias in pretrained language models,” *Available at SSRN 4754678*.
- SIMS, C. A. (2003): “Implications of rational inattention,” *Journal of monetary Economics*, 50, 665–690.
- THESMAR, D., AND E. VERNER (2025): “Beliefs and Stock Market Fluctuations: New Evidence from the Past Seven Decades,” *Available at SSRN*.
- 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

A Average Market Timing Alphas

This appendix reports the average market timing alphas summarized in the text. Each entry is the average alpha across the four annually rebalanced portfolios that begin in each quarter of the calendar year, estimated over the testing sample 2005:Q1-2023:Q4. The credible sets are constructed as described in Appendix B.2.2, which details how the average alphas are constructed.

B Construction of the credible sets

Credible sets appear in three places: the revision regressions in Table 3, the average market-timing alpha in Tables A.1 and A.2, and the subjective-versus-objective risk-return ratio in Table 4. All are flat-prior Bayesian posterior intervals of linear regressions. Section B.1 treats Table 3, whose coefficients each come from a single regression. Section B.2 treats the remaining tables, whose credible sets combine more than one regression whose estimates are correlated and are drawn jointly.

B.1 Credible sets for the revision regressions

Each coefficient reported in Table 3 comes from a single ordinary least squares regression. Under a flat prior the posterior of a coefficient is Student- t , centered at the OLS estimate, scaled by its standard error, with degrees of freedom equal to the residual degrees of freedom of that regression. The 90% credible set is the range from the 5th to the 95th percentile of this posterior. Because each estimate comes from a single equation, no cross-equation correlation enters. The joint construction in Section B.2 draws this same flat-prior posterior across equations to capture cross-equation correlation, and reduces to the single-equation case here when that correlation is zero.

B.2 Credible sets with cross-equation correlation

The credible sets for Tables 4, A.1, and A.2 each combine more than one regression whose coefficient estimates are correlated. We draw the correlated estimates jointly from their flat-prior posterior and read the credible set off the induced distribution. We first develop the

Table A.1: 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	0.045	0.649				
90% Credible Set	[0.02,0.07]	[0.54,0.76]				
FI	-0.002	0.993				
90% Credible Set	[-0.01,0.00]	[0.97,1.02]				
CFO	-0.003	0.988				
90% Credible Set	[-0.01,0.00]	[0.95,1.03]				
Livingston	0.000	1.000				
90% Credible Set	[0.00,0.00]	[1.00,1.00]				
Perfect Foresight	0.074	0.510				
90% Credible Set	[0.05,0.10]	[0.41,0.61]				
(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	0.049	0.546	0.553	-0.033	0.036	0.087
90% Credible Set	[0.03,0.07]	[0.45,0.64]	[0.35,0.75]	[-0.20,0.14]	[-0.22,0.29]	[-0.17,0.35]
FI	-0.003	1.001	-0.012	-0.033	-0.005	0.092
90% Credible Set	[-0.01,0.00]	[0.97,1.03]	[-0.09,0.06]	[-0.08,0.01]	[-0.07,0.06]	[0.01,0.17]
CFO	0.002	0.966	0.003	-0.034	-0.114	0.100
90% Credible Set	[-0.01,0.01]	[0.93,1.00]	[-0.09,0.10]	[-0.10,0.03]	[-0.20,-0.02]	[-0.01,0.21]
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	0.065	0.488	0.477	-0.059	0.248	0.132
90% Credible Set	[0.04,0.09]	[0.40,0.58]	[0.28,0.67]	[-0.22,0.10]	[0.01,0.48]	[-0.12,0.39]

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. Each reported coefficient is the average across the four annually-rebalanced portfolios that begin in successive quarters of the calendar year, each estimated on its own non-overlapping annual observations (nineteen per portfolio); the non-market factors are cumulated over the same forward four-quarter holding window as the strategy return. The Livingston survey, conducted semiannually, forms two such portfolios; the other forecasts form four. Each strategy is evaluated only over the quarters in which its own forecast exists. Panel (b) adds size (SMB), value (HML), profitability (RMW), and investment (CMA) factors. 90% credible sets are flat-prior Student- t intervals for the average across the four portfolios, built from a covariance matrix that includes the cross-equation correlation induced by the portfolios' overlapping four-quarter holding periods. Appendix B.2.2 gives the construction.

Table A.2: 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	0.070	0.445				
90% Credible Set	[0.03,0.11]	[0.27,0.62]				
FI	-0.002	0.993				
90% Credible Set	[-0.01,0.00]	[0.97,1.02]				
CFO	-0.003	0.988				
90% Credible Set	[-0.01,0.00]	[0.95,1.03]				
Livingston	0.000	1.000				
90% Credible Set	[0.00,0.00]	[1.00,1.00]				
Perfect Foresight	0.142	0.060				
90% Credible Set	[0.10,0.18]	[-0.13,0.25]				
(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	0.085	0.238	0.860	-0.087	-0.083	0.326
90% Credible Set	[0.05,0.12]	[0.08,0.39]	[0.57,1.15]	[-0.35,0.18]	[-0.49,0.32]	[-0.08,0.73]
FI	-0.003	1.001	-0.012	-0.033	-0.005	0.092
90% Credible Set	[-0.01,0.00]	[0.97,1.03]	[-0.09,0.06]	[-0.08,0.01]	[-0.07,0.06]	[0.01,0.17]
CFO	0.002	0.966	0.003	-0.034	-0.114	0.100
90% Credible Set	[-0.01,0.01]	[0.93,1.00]	[-0.09,0.10]	[-0.10,0.03]	[-0.20,-0.02]	[-0.01,0.21]
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	0.124	0.012	0.924	-0.106	0.469	0.289
90% Credible Set	[0.08,0.16]	[-0.16,0.19]	[0.55,1.30]	[-0.41,0.19]	[0.02,0.92]	[-0.19,0.77]

Notes: See Table A.1. 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. Each reported coefficient is the average across the four annually-rebalanced portfolios that begin in successive quarters of the calendar year, each estimated on its own non-overlapping annual observations (nineteen per portfolio); the non-market factors are cumulated over the same forward four-quarter holding window as the strategy return. The Livingston survey, conducted semiannually, forms two such portfolios; the other forecasts form four. Each strategy is evaluated only over the quarters in which its own forecast exists.

construction for the risk-return ratio and then apply it to the average market-timing alpha.

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

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

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

In the subjective risk-return trade-off regression, $y_t^{\mathbb{F}}$ represents survey-based return expectations, denoted $\mathbb{F}_t^s[r_{t,t+h}]$, while in the objective risk-return trade-off $y_t^{\mathbb{E}}$ represents machine-based return expectations, $\mathbb{E}_t[r_{t,t+h}]$. The regressor vector for the subjective regression is $X_t^{\mathbb{F}} = [1, \tilde{\mathbb{V}}_t[r_{t,t+h}]]'$, an intercept and subjective risk; for the objective regression $X_t^{\mathbb{E}} = [1, \mathbb{V}_t[r_{t,t+h}]]'$, an intercept and objective risk. The parameter vectors are $\beta_{\mathbb{F}} = [\beta_{\mathbb{F},0}, \beta_{\mathbb{F},1}]'$ and $\beta_{\mathbb{E}} = [\beta_{\mathbb{E},0}, \beta_{\mathbb{E},1}]'$.

Each regression has $k = 2$ coefficients. Stacking the $T_{\mathbb{F}}$ usable quarters of the subjective regression, the outcome $y^{\mathbb{F}}$ is $T_{\mathbb{F}} \times 1$ and the regressor matrix $X^{\mathbb{F}}$ is $T_{\mathbb{F}} \times 2$, a column of ones beside the subjective-risk column, so each row $X_t^{\mathbb{F}} = [1, \tilde{\mathbb{V}}_t]'$ is 2×1 , the coefficient vector $\beta_{\mathbb{F}}$ is 2×1 , and the posterior scale matrix $\hat{\sigma}_{\mathbb{F}}^2(X^{\mathbb{F}'}X^{\mathbb{F}})^{-1}$ is 2×2 . The objective regression has the same shape, with $y^{\mathbb{E}}$ of size $T_{\mathbb{E}} \times 1$ and $X^{\mathbb{E}}$ of size $T_{\mathbb{E}} \times 2$; the two samples need not share the same rows, since a quarter can be usable for one regression and not the other.

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 $T - k$ 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 ordinary least-squares estimate of the residual variance.¹ See pages 354–357, result (b) of Proposition 12.3. Specifically,

$$\begin{aligned} \beta_{\mathbb{F}} | y^{\mathbb{F}}, X^{\mathbb{F}} &\sim \mathcal{T}_{T_{\mathbb{F}}} \left(\hat{\beta}_{\mathbb{F}}, \hat{\sigma}_{\mathbb{F}}^2(X^{\mathbb{F}'}X^{\mathbb{F}})^{-1} \right) \\ \beta_{\mathbb{E}} | y^{\mathbb{E}}, X^{\mathbb{E}} &\sim \mathcal{T}_{T_{\mathbb{E}}} \left(\hat{\beta}_{\mathbb{E}}, \hat{\sigma}_{\mathbb{E}}^2(X^{\mathbb{E}'}X^{\mathbb{E}})^{-1} \right) \end{aligned}$$

where $\mathcal{T}_T(\mu, \Sigma)$ denotes a k -dimensional t -distribution with $T - k$ degrees of freedom, mean vector μ , and scale matrix Σ . The OLS estimator is $\hat{\beta} = (X'X)^{-1}X'y$, and the estimated

¹A sequence of proper priors that converges to the usual improper prior $p(\beta, \sigma^2) \propto 1/\sigma^2$ is $\beta | \sigma^2 \sim \mathcal{N}(0, \sigma^2 \tau^2 I_k)$ with $\tau^2 \rightarrow \infty$ and $\sigma^2 \sim \text{IG}(\varepsilon, \varepsilon)$ with $\varepsilon \rightarrow 0$; in the joint limit the prior converges to $p(\beta | \sigma^2) \propto 1$ and $p(\sigma^2) \propto 1/\sigma^2$, the flat prior on β together with the Jeffreys prior for σ^2 .

residual variance is $\hat{\sigma}^2 = \frac{1}{T-k} \sum_{t=1}^T (y_t - X_t \hat{\beta})^2$, with $k = 2$ the number of parameters in each regression.

Simulation Procedure. For $m = 1, \dots, M$ posterior draws, with $M = 100,000$ for our baseline results:

1. draw a realization of the subjective regression coefficients $\beta_{\mathbb{F}}^{(m)}$ from the multivariate t -distribution $\mathcal{T}_{T_{\mathbb{F}}}(\hat{\beta}_{\mathbb{F}}, \hat{\sigma}_{\mathbb{F}}^2 (X^{\mathbb{F}'} X^{\mathbb{F}})^{-1})$;
2. draw a realization of the objective regression coefficients $\beta_{\mathbb{E}}^{(m)}$ from $\mathcal{T}_{T_{\mathbb{E}}}(\hat{\beta}_{\mathbb{E}}, \hat{\sigma}_{\mathbb{E}}^2 (X^{\mathbb{E}'} X^{\mathbb{E}})^{-1})$;
3. compute the ratio of interest, $\theta^{(m)} = \beta_{\mathbb{F},1}^{(m)} / \beta_{\mathbb{E},1}^{(m)}$.

From the simulated posterior for θ we report two quantities: the 90% credible set, the interval between the 5th and 95th percentiles of the draws $\theta^{(m)}$, and the posterior probability $P(\theta < 1)$, the fraction of draws for which $\theta^{(m)} < 1$.

B.2.1 Treating the ratio as a joint system

Section B.2 draws $\beta_{\mathbb{F},1}$ and $\beta_{\mathbb{E},1}$ from separate posteriors, which imposes a correlation of zero between them. The two regressions are estimated on the same 2005:Q1–2023:Q4 sample, and in Panel D on the identical regressor, so that independence is unlikely to hold. We therefore treat the two regressions as a joint system, so the ratio’s uncertainty reflects however correlated the two estimates are, following the seemingly-unrelated-regressions framework of Zellner (1962).

Setup. The two regressions are those of Section B.2; $X^{\mathbb{F}}$ is $T_{\mathbb{F}} \times 2$ and $X^{\mathbb{E}}$ is $T_{\mathbb{E}} \times 2$, an intercept column and a risk column each. Within each equation the errors are serially uncorrelated, as the flat-prior sets of Section B.2 assume. Across the two equations in the same quarter the errors are allowed to be correlated, $\text{Cov}(\epsilon_t^{\mathbb{F}}, \epsilon_t^{\mathbb{E}}) = \sigma_{\mathbb{FE}}$, the piece the independent draws set to zero; across different quarters they are uncorrelated.

Step 1: fit both regressions. By ordinary least squares, $\hat{\beta}_{\mathbb{F}} = (X^{\mathbb{F}'} X^{\mathbb{F}})^{-1} X^{\mathbb{F}'} y^{\mathbb{F}}$ and likewise for $\mathbb{E}$. The point estimates, and the point ratio $\hat{\theta} = \hat{\beta}_{\mathbb{F},1} / \hat{\beta}_{\mathbb{E},1}$, are exactly those in Table 4; only the uncertainty around them changes. Retain $A_{\mathbb{F}} = (X^{\mathbb{F}'} X^{\mathbb{F}})^{-1}$ and $A_{\mathbb{E}} = (X^{\mathbb{E}'} X^{\mathbb{E}})^{-1}$.

Step 2: residuals. Keep the residuals $u_t^{\mathbb{F}} = y_t^{\mathbb{F}} - X_t^{\mathbb{F}'} \hat{\beta}_{\mathbb{F}}$ and $u_t^{\mathbb{E}} = y_t^{\mathbb{E}} - X_t^{\mathbb{E}'} \hat{\beta}_{\mathbb{E}}$, and let $\mathcal{T}$ be the set of quarters in which both regressions have a valid residual, of size T .

Step 3: cross-equation covariance. Over the overlap $\mathcal{T}$,

$$\hat{\sigma}_{\mathbb{F}\mathbb{E}} = \frac{1}{T} \sum_{t \in \mathcal{T}} u_t^{\mathbb{F}} u_t^{\mathbb{E}}.$$

Step 4: joint covariance of the two slopes. Each estimation error is a linear function of the errors, $\hat{\beta}_{\mathbb{F}} - \beta_{\mathbb{F}} = A_{\mathbb{F}} X^{\mathbb{F}'} \epsilon^{\mathbb{F}}$, so with $\text{Cov}(\epsilon^{\mathbb{F}}, \epsilon^{\mathbb{E}}) = \sigma_{\mathbb{F}\mathbb{E}} I$ over $\mathcal{T}$,

$$\text{Cov}(\hat{\beta}_{\mathbb{F}}, \hat{\beta}_{\mathbb{E}}) = \sigma_{\mathbb{F}\mathbb{E}} A_{\mathbb{F}} (X^{\mathbb{F}'} X^{\mathbb{E}}) A_{\mathbb{E}},$$

and $\text{Cov}(\hat{\beta}_{\mathbb{F},1}, \hat{\beta}_{\mathbb{E},1})$ is its $(2,2)$ entry. In Panel D, where $X^{\mathbb{F}} = X^{\mathbb{E}} = X$, this collapses to $\sigma_{\mathbb{F}\mathbb{E}} (X'X)^{-1}$. With the two slope variances already used in Section B.2, $\widehat{\text{Var}}(\hat{\beta}_{\mathbb{F},1})$ and $\widehat{\text{Var}}(\hat{\beta}_{\mathbb{E},1})$ (the slope entries of $\hat{\sigma}_{\mathbb{F}}^2 A_{\mathbb{F}}$ and $\hat{\sigma}_{\mathbb{E}}^2 A_{\mathbb{E}}$), the 2×2 joint covariance of the two slopes is

$$\Sigma = \begin{pmatrix} \widehat{\text{Var}}(\hat{\beta}_{\mathbb{F},1}) & \widehat{\text{Cov}}(\hat{\beta}_{\mathbb{F},1}, \hat{\beta}_{\mathbb{E},1}) \\ \widehat{\text{Cov}}(\hat{\beta}_{\mathbb{F},1}, \hat{\beta}_{\mathbb{E},1}) & \widehat{\text{Var}}(\hat{\beta}_{\mathbb{E},1}) \end{pmatrix}.$$

In dimensions, $A_{\mathbb{F}}$ and $A_{\mathbb{E}}$ are 2×2 , and the cross-product $X^{\mathbb{F}'} X^{\mathbb{E}}$ is 2×2 as well: the overlap sample $\mathcal{T}$ is the inner dimension of $X^{\mathbb{F}'}$ and $X^{\mathbb{E}}$ and is summed over, not left in the result. The sandwich therefore returns the 2×2 matrix $\text{Cov}(\hat{\beta}_{\mathbb{F}}, \hat{\beta}_{\mathbb{E}})$, whose lower-right entry is the scalar slope covariance $\text{Cov}(\hat{\beta}_{\mathbb{F},1}, \hat{\beta}_{\mathbb{E},1})$ that enters Σ .

Step 5: draw jointly. Take the 2×2 lower-triangular Cholesky factor $L = \begin{pmatrix} L_{11} & 0 \\ L_{21} & L_{22} \end{pmatrix}$ of Σ , with

$$L_{11} = \sqrt{\Sigma_{11}}, \quad L_{21} = \frac{\Sigma_{21}}{L_{11}}, \quad L_{22} = \sqrt{\Sigma_{22} - L_{21}^2},$$

so that $LL' = \Sigma$; if $\widehat{\text{Cov}}(\hat{\beta}_{\mathbb{F},1}, \hat{\beta}_{\mathbb{E},1}) = 0$ then $L_{21} = 0$ and the draws reduce to the independent ones of Section B.2. For $m = 1, \dots, M$: draw $z_{1,m}, z_{2,m} \sim \mathcal{N}(0, 1)$ and form $w_m = Lz_m$; draw one shared $\chi_{\nu,m}^2$ with $\nu = T - 2$ degrees of freedom and scale $v_m = \sqrt{\nu / \chi_{\nu,m}^2} w_m$ to make the pair Student- t ; set $(\beta_{\mathbb{F},1}^{(m)}, \beta_{\mathbb{E},1}^{(m)}) = (\hat{\beta}_{\mathbb{F},1}, \hat{\beta}_{\mathbb{E},1}) + v_m$; and compute $\theta^{(m)} = \beta_{\mathbb{F},1}^{(m)} / \beta_{\mathbb{E},1}^{(m)}$. As in Section B.2 we report the 5th and 95th percentiles of $\theta^{(m)}$ and the share of draws below one; only the mechanism generating the pairs has changed, from two independent columns of draws to one jointly-drawn pair.

B.2.2 Credible sets for the average alpha

The long-only and long-short alpha tables (Tables A.1 and A.2) each report a single alpha that averages four annually rebalanced strategies, one formed in each quarter of the calendar year and held for four quarters. This section explains how that average is formed and how its credible set is built. The four strategies overlap in calendar time, so their alpha estimates are correlated. We first form the average from four non-overlapping subsamples (Steps 1 and 2), then widen the set to account for the overlap (Steps 3 to 5).

A portfolio formed at the end of quarter τ holds the position implied by the forecast made at τ over the four quarters $\tau + 1, \dots, \tau + 4$, so the dependent variable $r_{s,t,t+4} - r_{f,t}$ is a four-quarter log return and consecutive quarterly observations overlap in three of their four quarters. We split the overlapping observations by formation quarter into four subsamples: subsample $j = 1$ collects the observations formed in the first quarter of each year, $j = 2$ those formed in the second, $j = 3$ the third, and $j = 4$ the fourth. Each subsample has one observation per year, so within a subsample the four-quarter windows do not overlap. There are $T = 19$ annual observations in each subsample; Livingston, observed semiannually, forms two of the four.

Step 1: Fit each subsample. On subsample j we estimate by ordinary least squares

$$r_{s,t,t+4} - r_{f,t} = \alpha_j + \beta_j' F_t + \varepsilon_{j,t}, \quad t \in T_j,$$

where F_t is the factor vector, cumulated in logs over the same four-quarter window as the strategy return: under the CAPM F_t is the scalar market excess return, and under the five-factor model F_t is the 5×1 vector of Fama-French factors. Write $b_j = (\alpha_j, \beta_j')'$ for the coefficient vector and K for its length, so $K = 2$ (an intercept and the market loading) under the CAPM and $K = 6$ (an intercept and five loadings) under the five-factor model. Stacking the $T = 19$ annual observations of subsample j , the outcome y_j is $T \times 1$ and the regressor matrix X_j is $T \times K$, with t -th row $[1, F_t']$; the least-squares estimate $\hat{b}_j = (X_j' X_j)^{-1} X_j' y_j$ is $K \times 1$ and $(X_j' X_j)^{-1}$ is $K \times K$. The alpha the tables report is the simple average across the four formation quarters,

$$\bar{b} = \frac{1}{4} \sum_{j=1}^4 b_j, \quad \bar{\alpha} = \frac{1}{4} \sum_j \alpha_j, \quad \bar{\beta} = \frac{1}{4} \sum_j \beta_j,$$

a $K \times 1$ vector whose first entry is the reported average alpha. Averaging the four is what lets each table report one alpha per strategy rather than four formation-quarter estimates.

Step 2: The credible set that treats the four as independent. Each subsample has the flat-prior Student- t posterior of Section B.2, now K -dimensional. We draw from it by the standard reparameterization. Let $\hat{s}_j^2$ be the (scalar) residual variance on subsample j and let L_j be the $K \times K$ lower-triangular Cholesky factor of the $K \times K$ matrix $\hat{s}_j^2(X_j'X_j)^{-1}$, the unique lower-triangular matrix with $L_jL_j' = \hat{s}_j^2(X_j'X_j)^{-1}$. A draw is

$$b_j^{(m)} = \hat{b}_j + \sqrt{\frac{T-K}{\chi_{j,m}^2}} L_j z_{j,m}, \quad z_{j,m} \sim \mathcal{N}(0, I_K), \quad \chi_{j,m}^2 \sim \chi_{T-K}^2,$$

where $z_{j,m}$ is a $K \times 1$ standard normal vector, I_K the $K \times K$ identity, and the scalar χ_{T-K}^2 draw turns the Gaussian $L_j z_{j,m}$ into a Student- t , delivering the whole $K \times 1$ vector, alpha and factor loadings together. On each draw we average the four, $\bar{b}^{(m)} = \frac{1}{4} \sum_j b_j^{(m)}$; repeating $M = 100,000$ times gives the posterior of each averaged coefficient, whose 5th and 95th percentiles are the reported set. Because the four subsamples share no observations, treating the draws as independent gives the average an alpha variance

$$\widehat{\text{Var}}(\bar{\alpha}) = \frac{1}{16} \sum_{j=1}^4 \widehat{\text{Var}}(\hat{\alpha}_j),$$

a scalar, where $\widehat{\text{Var}}(\hat{\alpha}_j)$ is the (1,1) entry of the $K \times K$ matrix $\hat{s}_j^2(X_j'X_j)^{-1}$. Steps 3 to 5 widen this independent set to the one reported in Tables A.1 and A.2.

Step 3: Estimate the cross-equation residual covariances. The four subsamples use different formation dates, but their four-quarter holding periods overlap in calendar time. To correct for this overlap, we draw the four jointly, starting from the covariance between their errors. Align the residuals $\hat{\varepsilon}_{j,Y}$ of the four subsample regressions by calendar year Y . Two subsamples' holding periods overlap only at two horizons, the same year and one year apart, so we estimate the scalar cross-equation error covariances at those two horizons,

$$\hat{\sigma}_{jk}^{(0)} = \frac{1}{T-K} \sum_Y \hat{\varepsilon}_{j,Y} \hat{\varepsilon}_{k,Y}, \quad \hat{\sigma}_{jk}^{(1)} = \frac{1}{T-K} \sum_Y \hat{\varepsilon}_{j,Y} \hat{\varepsilon}_{k,Y-1}.$$

The first asks whether subsamples j and k tend to have large residuals in the same year; the second asks the same for years one apart. If the subsamples were in fact uncorrelated, both

would come out near zero and the rest of the adjustment would vanish.

Step 4: Carry the residual covariance to the coefficients. Substituting the model into the least-squares formula gives $\hat{b}_j - b_j = (X_j' X_j)^{-1} X_j' \varepsilon_j$, a $K \times 1$ vector. The covariance between two coefficient vectors is therefore a $K \times K$ matrix,

$$\widehat{\text{Cov}}(\hat{b}_j, \hat{b}_k) = (X_j' X_j)^{-1} M_{jk} (X_k' X_k)^{-1},$$

$$M_{jk} = \hat{\sigma}_{jk}^{(0)} \sum_Y X_{j,Y} X_{k,Y}' + \hat{\sigma}_{jk}^{(1)} \sum_Y X_{j,Y} X_{k,Y-1}' + \hat{\sigma}_{kj}^{(1)} \sum_Y X_{j,Y} X_{k,Y+1}',$$

where $X_{j,Y}$ is subsample j 's $K \times 1$ regressor row for year Y (written as a column). Each outer product $X_{j,Y} X_{k,Y}'$ is $K \times K$, so M_{jk} is $K \times K$ and, flanked by the two $K \times K$ matrices $(X_j' X_j)^{-1}$ and $(X_k' X_k)^{-1}$, $\widehat{\text{Cov}}(\hat{b}_j, \hat{b}_k)$ is $K \times K$.

Step 5: Assemble the average and its credible set. The averaged coefficient vector then has the $K \times K$ covariance

$$\widehat{\text{Var}}(\bar{b}) = \frac{1}{16} \sum_{j=1}^4 \sum_{k=1}^4 \widehat{\text{Cov}}(\hat{b}_j, \hat{b}_k),$$

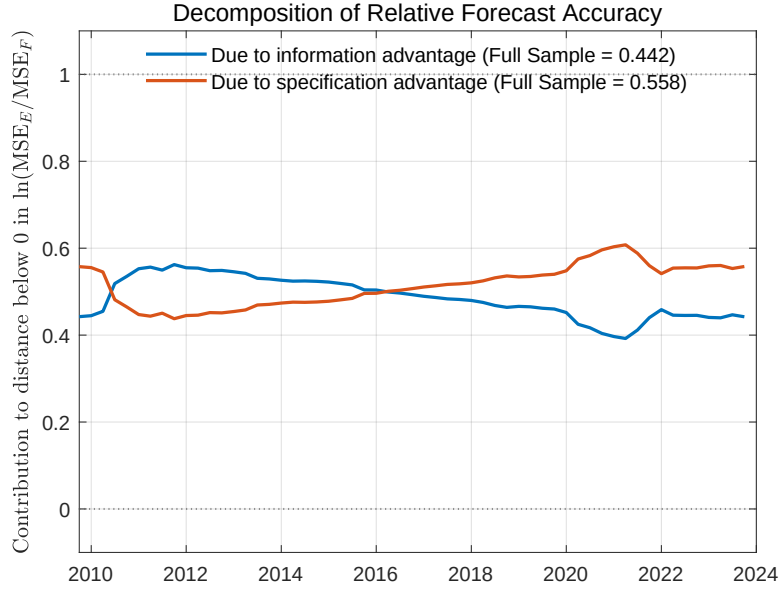
which sums all sixteen $K \times K$ blocks rather than only the four diagonal ones. Setting the $j \neq k$ terms to zero recovers the independent variance of Step 2. The credible set for each averaged coefficient is the flat-prior Student- t interval of Section B.2 with $T - K$ degrees of freedom, built from the corresponding diagonal entry of $\widehat{\text{Var}}(\bar{b})$; the average alpha uses its $(1, 1)$ entry. T

Tables A.1 and A.2 report the cross-equation-adjusted credible sets.

C Additional Results

Decomposition of Relative Forecast Accuracy

Figure C.1: Decomposition of Relative Forecast Accuracy



Notes: The figure shows how much of the machine's improvement in forecast accuracy over the IBES survey comes from using more information versus specification. Let $\mathbb{E}$ denote the unrestricted machine forecast of 1-year earnings growth, $\mathbb{E}_1$ the restricted machine forecast using only the equity cash-flow theme as input, and $\mathbb{F}$ the IBES survey forecast. The total log MSE ratio decomposes as $\ln[\text{MSE}(\mathbb{E})/\text{MSE}(\mathbb{F})] = \ln[\text{MSE}(\mathbb{E})/\text{MSE}(\mathbb{E}_1)] + \ln[\text{MSE}(\mathbb{E}_1)/\text{MSE}(\mathbb{F})]$. Dividing each term by the total yields shares that sum to one. The share due to information advantage, $\ln[\text{MSE}(\mathbb{E})/\text{MSE}(\mathbb{E}_1)]/\ln[\text{MSE}(\mathbb{E})/\text{MSE}(\mathbb{F})]$, captures the contribution of the broader input set (holding architecture fixed at the LSTM); the share due to specification advantage, $\ln[\text{MSE}(\mathbb{E}_1)/\text{MSE}(\mathbb{F})]/\ln[\text{MSE}(\mathbb{E})/\text{MSE}(\mathbb{F})]$, captures the contribution of the LSTM architecture (holding the input set fixed at the equity cash-flow theme). MSEs are computed over expanding windows starting in 2005:Q1 with a five-year minimum.

Machine Forecast Errors and Past Revisions

Figure C.2 examines whether the machine forecasts display the predictable errors that characterize the survey forecasts. For both earnings growth and stock returns, past forecast revisions do not predict subsequent machine forecast errors, indicating that the machine incorporates the information in its revisions efficiently.

Figure C.2: Machine Forecast Errors and Past Revisions

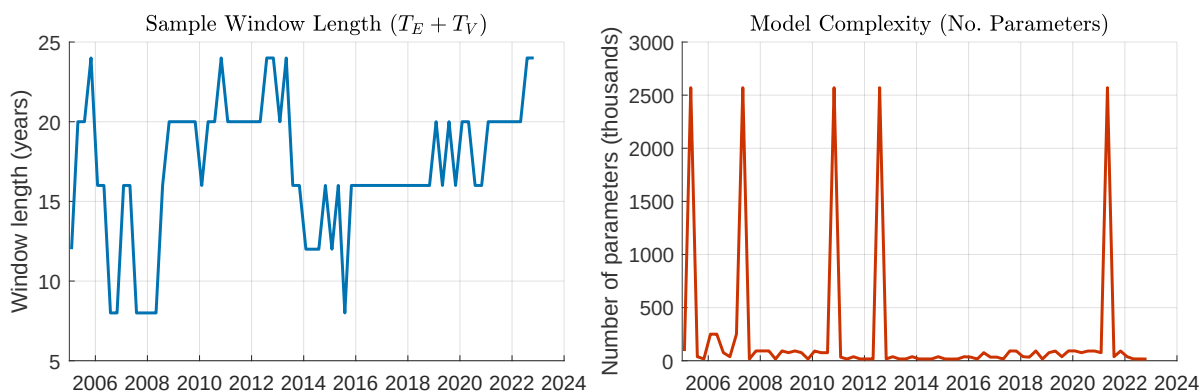


Notes: Each panel plots the estimated coefficient on the machine forecast revision at each forecast date. A bar appears only where that coefficient is nonzero. The left panel uses one-year stock returns and the right panel one-year earnings growth. At each forecast date t , the machine forecast error is regressed on the machine forecast revision and all real-time predictors used by the corresponding LSTM, using only information available at date t and the period-specific estimation and validation windows selected by the LSTM procedure. The penalty parameter is chosen to minimize validation-sample mean squared error within that real-time window choice.

Machine Forecast: Training Window and Model Complexity

Figure C.3 plots the total training window $T_E + T_V$ chosen by the machine each period (left panel) and the implied number of LSTM parameters under the pyramid rule (right panel), for the earnings growth forecast. During the second half of the GFC and the subsequent rebound, the machine extends its window to cover the dot-com recession and selects a more complex architecture. This allows the machine to train on longer historical data and increase its model flexibility to better extract information from this long-term history.

Figure C.3: Machine Forecast: Training Window and Model Complexity



Notes: Left panel plots $T_E + T_V$, the total training and validation window chosen by the machine each period for the earnings growth forecast. Right panel plots the implied number of LSTM parameters under the pyramid rule. During the second half of the GFC and the subsequent rebound, the machine extends its training window to cover the dot-com recession and selects a more complex architecture, consistent with the LSTM tapping long memory around turning points.

D 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 $\ln\text{CRSPD} = 12\ln(1 + \text{VWRETD})$. 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, $\log\left(\frac{P_{t+12}^{\text{S\&P}}}{P_t^{\text{S\&P}}}\right)$. 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 methodology used in 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. The divisor incorporates changes in shares outstanding and changes in mechanical index construction effects arising from changes in constituents in order to give an index-adjusted scale that represents the true economic earnings. This ensures that earnings is comparable across time even when constituent firms change. 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 in rare occasions 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 for aggregate earnings h -months-ahead as

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

where E is aggregate S&P 500 earnings, $\mathbb{F}_{med,t}[EPS_{i,t+h}]$ 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. The divisor incorporates changes in shares outstanding and changes in mechanical index construction effects arising from changes in constituents in order to give an index-adjusted scale that represents the true economic earnings. This ensures that earnings is comparable across time even when constituent firms change. 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 Ω_{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 in rare occasions 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 De La O and Myers (2021):

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

where e_t is log earnings for S&P 500 at time t calculated as $e_t = \ln(EPS_t \cdot Divisor_t)$, 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 on 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 Gallup/UBS 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 Gallup/UBS 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 (SPIF_12M_t) and 0-month level nowcasts (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 (spi12_t) and 6-month (spi6_t) level forecast of the S&P 500 index

$$\mathbb{F}_t^{(Liv)} \left[\frac{P_{t+12}}{P_t} \right] \approx \begin{cases} \frac{\mathbb{F}_t^{(Liv)}[P_{t+12}]}{\mathbb{F}_t^{(Liv)}[P_t]} = \frac{\text{SPIF_12M}_t}{\text{SPIF_ZM}_t} & \text{if } t \geq 1992M6 \\ \left(\frac{\mathbb{F}_t^{(Liv)}[P_{t+12}]}{\mathbb{F}_t^{(Liv)}[P_{t+6}]} \right)^2 = \left(\frac{\text{spi12}_t}{\text{spi6}_t} \right)^2 & \text{if } t < 1992M6 \end{cases} \quad (\text{D.1})$$

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:

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

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).

Value Line Expected Returns Value Line is an independent equity research firm that has published forecasts of firm-level fundamentals and stock prices since 1956. We use the aggregate VL expected nominal return series constructed by Thesmar and Verner (2025), who build a long firm-level panel of VL expectations covering 1956 to 2024 by digitizing physical volumes and microfilm records for the 1956–1989 period and combining these with VL’s digital database for the 1987–2024 period. Their dataset covers 92,117 firm-year observations across 38,696 investment analysis reports, with VL firms accounting for over 80% of CRSP market capitalization in most years. For each firm, VL provides forecasts at the three-to-five year horizon along with a stock price target at the same horizon. Aggregate VL expected nominal returns, denoted $F_{t \rightarrow t+4}^{VL}$, are computed by solving forward from the price target, the current price, and projected earnings per share, and are annualized over the three-to-five year horizon. We use the quarterly aggregate series available from 1987:Q3 onward as our $h = 16$ survey forecast Thesmar and Verner (2025). We thank David Thesmar for generously providing their Value Line expected returns data. For details on data construction, see Thesmar and Verner (2025). We obtained the data on January 6, 2026.

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, $\mathbb{F}_t^{\text{CFO}}[r_{t,t+4}]$, 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, $P_{t,t+4}^{\text{SOC}}$, by:

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

The final imputed SOC point forecast is constructed as $\mathbb{F}_t^{\text{SOC}}[r_{t,t+4}] = \hat{\beta}_0 + \hat{\beta}_1 P_{t,t+4}^{\text{SOC}} + D_t/P_t$. 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:Q3 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 $\mathbb{F}_t^{\text{SOC}}[r_{t,t+4}]$ 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 $P_{t,t+4}^{\text{CB}}$ 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.

Table D.1: 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.

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

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

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 variance of the latent innovation error. Observation equation takes the form $X_t = D + ZS_t + Uv_t$

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

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 a common variance, so that the observation-noise matrix U is a scalar multiple of the identity; the same observation-noise variance applies to every survey. 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:

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

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 and CB, 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 D 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 1989:10-2023:12 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 1999:04-2023:12 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 financial market returns, over a 12-month horizon. The data has been downloaded on September 25, 2024 from the author's website: <https://www.sydneyludvigson.com/s/>

Factor Data

This section describes the data sources for the factor returns used in Table A.2. 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 nine small-cap portfolios minus the average return on nine large-cap portfolios, formed from the intersections of three independent 2x3 sorts on size and book-to-market, size and operating profitability, and size and investment. 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 5 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* from January 1984 to June 2024. The current vintage of the results reported here is based on a randomly selected subsample 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.

E 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, 3, 4, 6, 12, and 24 month-ahead federal funds futures (FFF) contract rate and changes in the 1, 2, 3, 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 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 E.1 gives the complete list of real-time macro variables. Included in the table is the first available vintages for each variable that has multiple vintages. We do not include the last vintage because most variables have vintages through the present.³ Table E.1 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 E.1: 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

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

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

Table E.1 (Cont'd)

No.	Short Name	Source	Tran	Description	First Vintage
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	OPHMQVD	Philly Fed	Δln	Output per hour - Business sector	1998:Q4
8	NDPIQVQD	Philly Fed	Δln	Nom. disposable personal income	1965:Q4
9	NOUTPUTQVQD	Philly Fed	Δln	Nom. GNP/GDP	1965:Q4
10	NPIQVQD	Philly Fed	Δln	Nom. personal income	1965:Q4
11	NPSAVQVQD	Philly Fed	Δlv	Nom. personal saving	1965:Q4
12	OLIQVQD	Philly Fed	Δln	Other labor income	1965:Q4
13	PINTIQVQD	Philly Fed	Δln	Personal interest income	1965:Q4
14	PINTPAIDQVQD	Philly Fed	Δln	Interest paid by consumers	1965:Q4
15	PROPIQVQD	Philly Fed	Δln	Proprietors' income	1965:Q4
16	PTAXQVQD	Philly Fed	Δln	Personal tax and nontax payments	1965:Q4
17	RATESAVQVQD	Philly Fed	Δlv	Personal saving rate	1965:Q4
18	RENTIQVQD	Philly Fed	Δlv	Rental income of persons	1965:Q4
19	ROUTPUTQVQD	Philly Fed	Δln	Real GNP/GDP	1965:Q4
20	SSCONTRIBQVQD	Philly Fed	Δln	Personal contributions for social insurance	1965:Q4
21	TRANPFQVQD	Philly Fed	Δln	Personal transfer payments to foreigners	1965:Q4
22	TRANRQVQD	Philly Fed	Δln	Transfer payments	1965:Q4
23	CUUR0000SA0E	BLS	$\Delta^2 ln$	Energy in U.S. city avg., all urban consumers, not seasonally adj	
Group 2: Employment					
24	EMPLOYMVMD	Philly Fed	Δln	Nonfarm payroll	1946:M12
25	HMVMD	Philly Fed	lv	Aggregate weekly hours - Total	1971:M9
26	HGMVMD	Philly Fed	lv	Agg. weekly hours - Goods-producing	1971:M9
27	HSMVMD	Philly Fed	lv	Agg. weekly hours - Service-producing	1971:M9
28	LFCMVMD	Philly Fed	Δln	Civilian labor force	1998:M11

Table E.1 (Cont'd)

No.	Short Name	Source	Tran	Description	First Vintage
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	RINVBFBMVQD	Philly Fed	Δln	Real gross private domestic inv. - Nonresidential	1965:Q4
36	RINVCHIMVQD	Philly Fed	Δlv	Real gross private domestic inv. - Change in private inventories	1965:Q4
37	RINVRESIDMVQD	Philly Fed	Δln	Real gross private domestic inv. - Residential	1965:Q4
38	CASESHILLER	S&P	Δln	Case-Shiller US National Home Price index/CPI	1987:M1
Group 4: Consumption					
39	NCONGMMVMD	Philly Fed	Δln	Nom. personal cons. exp. - Goods	2009:M8
40	NCONHHMMVMD	Philly Fed	Δln	Nom. hh. cons. exp.	2009:M8
41	NCONSHHMMVMD	Philly Fed	Δln	Nom. hh. cons. exp. - Services	2009:M8
42	NCONSNPMMVMD	Philly Fed	Δln	Nom. final cons. exp. of NPISH	2009:M8
43	RCONDMMVMD	Philly Fed	Δln	Real personal cons. exp. - Durables	1998:M11
44	RCONGMMVMD	Philly Fed	Δln	Real personal cons. exp. - Goods	2009:M8
45	RCONHHMMVMD	Philly Fed	Δln	Real hh. cons. exp.	2009:M8
46	RCONMMVMD	Philly Fed	Δln	Real personal cons. exp. - Total	1998:M11
47	RCONNDMVMD	Philly Fed	Δln	Real personal cons. exp. - Nondurables	1998:M11
48	RCONSHHMMVMD	Philly Fed	Δln	Real hh. cons. exp. - Services	2009:M8
49	RCONSMMVMD	Philly Fed	Δln	Real personal cons. exp. - Services	1998:M11
50	RCONSNPMMVMD	Philly Fed	Δln	Real final cons. exp. of NPISH	2009:M8
51	NCONGMVQD	Philly Fed	Δln	Nom. personal cons. exp. - Goods	2009:Q3
52	NCONHHMVQD	Philly Fed	Δln	Nom. hh. cons. exp.	2009:Q3
53	NCONSHHMVQD	Philly Fed	Δln	Nom. hh. cons. exp. - Services	2009:Q3

Table E.1 (Cont'd)

No.	Short Name	Source	Tran	Description	First Vintage
54	NCONSNPMVQD	Philly Fed	$\Delta \ln$	Nom. final cons. exp. of NPISH	2009:Q3
55	RCONDMVQD	Philly Fed	$\Delta \ln$	Real personal cons. exp. - Durable goods	1965:Q4
56	RCONGMVQD	Philly Fed	$\Delta \ln$	Real personal cons. exp. - Goods	2009:Q3
57	RCONHHMVQD	Philly Fed	$\Delta \ln$	Real hh. cons. exp.	2009:Q3
58	RCONMVQD	Philly Fed	$\Delta \ln$	Real personal cons. exp. - Total	1965:Q4
59	RCONNDMVQD	Philly Fed	$\Delta \ln$	Real personal cons. exp. - Nondurable goods	1965:Q4
60	RCONSHHMVQD	Philly Fed	$\Delta \ln$	Real hh. cons. exp. - Services	2009:Q3
61	RCONSMVQD	Philly Fed	$\Delta \ln$	Real personal cons. exp. - Services	1965:Q4
62	RCONSNPMVQD	Philly Fed	$\Delta \ln$	Real final cons. exp. of NPISH	2009:Q3
63	NCONQVQD	Philly Fed	$\Delta \ln$	Nom. personal cons. exp.	1965:Q4
Group 5: Prices					
64	PCONGMMVMD	Philly Fed	$\Delta^2 \ln$	Price index for personal cons. exp. - Goods	2009:M8
65	PCONHHMMVMD	Philly Fed	$\Delta^2 \ln$	Price index for hh. cons. exp.	2009:M8
66	PCONSHHMVMD	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

Table E.1 (Cont'd)

No.	Short Name	Source	Tran	Description	First Vintage
Group 6: Trade and Government					
81	REXMVQD	Philly Fed	$\Delta \ln$	Real exports of goods and services	1965:Q4
82	RGMVQD	Philly Fed	$\Delta \ln$	Real government cons. and gross inv. - Total	1965:Q4
83	RGFMVQD	Philly Fed	$\Delta \ln$	Real government cons. and gross inv. - Federal	1965:Q4
84	RGSLMVQD	Philly Fed	$\Delta \ln$	Real government cons. and gross. inv. - State and local	1965:Q4
85	RIMPMVQD	Philly Fed	$\Delta \ln$	Real imports of goods and services	1965:Q4
86	RNXMVQD	Philly Fed	$\Delta \ln$	Real net exports of goods and services	1965:Q4
Group 7: Money and Credit					
87	BASEBASAQVMD	Philly Fed	$\Delta^2 \ln$	Monetary base	1980:Q2
88	M1QVMD	Philly Fed	$\Delta^2 \ln$	M1 money stock	1965:Q4
89	M2QVMD	Philly Fed	$\Delta^2 \ln$	M2 money stock	1971:Q2
90	NBRBASAQVMD	Philly Fed	$\Delta \ln / \ln$	Nonborrowed reserves	1967:Q3
91	NBRECASAQVMD	Philly Fed	$\Delta \ln / \ln$	Nonborrowed reserves plus extended credit	1984:Q2
92	TRBASAQVMD	Philly Fed	$\Delta^2 \ln$	Total reserves	1967:Q3
93	DIVQVQD	Philly Fed	$\Delta \ln$	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 Kourtellis (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 E.2 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 E.2: List of Daily Financial Dataset Variables

No.	Short Name	Source	Tran	Description
Group 1: Commodities				
1	GSIJSPT	Data Stream	$\Delta \ln$	S&P GSCI Zinc Spot - PRICE INDEX
2	GSSBSPT	Data Stream	$\Delta \ln$	S&P GSCI Sugar Spot - PRICE INDEX
3	GSSOSPT	Data Stream	$\Delta \ln$	S&P GSCI Soybeans Spot - PRICE INDEX
4	GSSISPT	Data Stream	$\Delta \ln$	S&P GSCI Silver Spot - PRICE INDEX
5	GSIKSPT	Data Stream	$\Delta \ln$	S&P GSCI Nickel Spot - PRICE INDEX
6	GSLCSPT	Data Stream	$\Delta \ln$	S&P GSCI Live Cattle Spot - PRICE INDEX
7	GSLHSPT	Data Stream	$\Delta \ln$	S&P GSCI Lean Hogs Index Spot - PRICE INDEX
8	GSILSPT	Data Stream	$\Delta \ln$	S&P GSCI Lead Spot - PRICE INDEX
9	GSGCSPT	Data Stream	$\Delta \ln$	S&P GSCI Gold Spot - PRICE INDEX

Table E.2 (Cont'd)

No.	Short Name	Source	Tran	Description
10	GSCTSPT	Data Stream	$\Delta \ln$	S&P GSCI Cotton Spot - PRICE INDEX
11	GSKCSPT	Data Stream	$\Delta \ln$	S&P GSCI Coffee Spot - PRICE INDEX
12	GSCCSPT	Data Stream	$\Delta \ln$	S&P GSCI Cocoa Index Spot - PRICE INDEX
13	GSIASPT	Data Stream	$\Delta \ln$	S&P GSCI Aluminum Spot - PRICE INDEX
14	SGWTSPT	Data Stream	$\Delta \ln$	S&P GSCI All Wheat Spot - PRICE INDEX
15	EIAEBRT	Data Stream	$\Delta \ln$	Europe Brent Spot FOB U\$/BBL Daily
16	CRUDOIL	Data Stream	$\Delta \ln$	Crude Oil-WTI Spot Cushing U\$/BBL - MID PRICE
17	LTICASH	Data Stream	$\Delta \ln$	LME-Tin 99.85% Cash U\$/MT
18	CWFCS00	Data Stream	$\Delta \ln$	CBT-WHEAT COMPOSITE FUTURES CONT. - SETT. PRICE
19	CCFCS00	Data Stream	$\Delta \ln$	CBT-CORN COMP. CONTINUOUS - SETT. PRICE
20	CSYCS00	Data Stream	$\Delta \ln$	CBT-SOYBEANS COMP. CONT. - SETT. PRICE
21	NCTCS20	Data Stream	$\Delta \ln$	CSCE-COTTON #2 CONT.2ND FUT - SETT. PRICE
22	NSBCS00	Data Stream	$\Delta \ln$	CSCE-SUGAR #11 CONTINUOUS - SETT. PRICE
23	NKCCS00	Data Stream	$\Delta \ln$	CSCE-COFFEE C CONTINUOUS - SETT. PRICE
24	NCCCS00	Data Stream	$\Delta \ln$	CSCE-COCOA CONTINUOUS - SETT. PRICE
25	CZLCS00	Data Stream	$\Delta \ln$	ECBOT-SOYBEAN OIL CONTINUOUS - SETT. PRICE
26	COFC01	Data Stream	$\Delta \ln$	CBT-OATS COMP. TRc1 - SETT. PRICE
27	CLDCS00	Data Stream	$\Delta \ln$	CME-LIVE CATTLE COMP. CONTINUOUS - SETT. PRICE
28	CLGC01	Data Stream	$\Delta \ln$	CME-LEAN HOGS COMP. TRc1 - SETT. PRICE
29	NGCCS00	Data Stream	$\Delta \ln$	CMX-GOLD 100 OZ CONTINUOUS - SETT. PRICE
30	LAH3MTH	Data Stream	$\Delta \ln$	LME-Aluminium 99.7% 3 Months U\$/MT
31	LED3MTH	Data Stream	$\Delta \ln$	LME-Lead 3 Months U\$/MT
32	LNI3MTH	Data Stream	$\Delta \ln$	LME-Nickel 3 Months U\$/MT
33	LTi3MTH	Data Stream	$\Delta \ln$	LME-Tin 99.85% 3 Months U\$/MT
34	PLNYD	www.macrotrends.net	$\Delta \ln$	Platinum Cash Price (U\$ per troy ounce)
35	XPDD	www.macrotrends.net	$\Delta \ln$	Palladium (U\$ per troy ounce)

Table E.2 (Cont'd)

No.	Short Name	Source	Tran	Description
36	CUS2D	www.macrotrends.net	Δln	Corn Spot Price (US\$/Bushel)
37	SoybOil	www.macrotrends.net	Δln	Soybean Oil Price (US\$/Pound)
38	OATSD	www.macrotrends.net	Δln	Oat Spot Price (US\$/Bushel)
39	WTIOilFut	US EIA	Δln	Light Sweet Crude Oil Futures Price: 1St Expiring Contract Settlement (\$/Bbl)
Group 2: Equities				
40	S&PCOMP	Data Stream	Δln	S&P 500 COMPOSITE - PRICE INDEX
41	ISPCS00	Data Stream	Δln	CME-S&P 500 INDEX CONTINUOUS - SETT. PRICE
42	SP5EIND	Data Stream	Δln	S&P500 ES INDUSTRIALS - PRICE INDEX
43	DJINDUS	Data Stream	Δln	DOW JONES INDUSTRIALS - PRICE INDEX
44	CYMCS00	Data Stream	Δln	CBT-MINI DOW JONES CONTINUOUS - SETT. PRICE
45	NASCOMP	Data Stream	Δln	NASDAQ COMPOSITE - PRICE INDEX
46	NASA100	Data Stream	Δln	NASDAQ 100 - PRICE INDEX
47	CBOEVIX	Data Stream	lv	CBOE SPX VOLATILITY VIX (NEW) - PRICE INDEX
48	S&P500toVIX	Data Stream	Δln	S&P500/VIX
Group 3: Corporate Risk				
49	LIBOR	FRED	Δlv	Overnight London Interbank Offered Rate (%)
50	1MLIBOR	FRED	Δlv	1-Month London Interbank Offered Rate (%)
51	3MLIBOR	FRED	Δlv	3-Month London Interbank Offered Rate (%)
52	6MLIBOR	FRED	Δlv	6-Month London Interbank Offered Rate (%)
53	1YLIBOR	FRED	Δlv	One-Year London Interbank Offered Rate (%)
54	1MEuro-FF	FRED	lv	1-Month Eurodollar Deposits (London Bid) (% P.A.) minus Fed Funds
55	3MEuro-FF	FRED	lv	3-Month Eurodollar Deposits (London Bid) (% P.A.) minus Fed Funds
56	6MEuro-FF	FRED	lv	6-Month Eurodollar Deposits (London Bid) (% P.A.) minus Fed Funds

Table E.2 (Cont'd)

No.	Short Name	Source	Tran	Description
57	APFNF-AANF	Data Stream	<i>lv</i>	1-Month A2/P2/F2 Nonfinancial Commercial Paper (NCP) (% P. A.) minus 1-Month Aa NCP (% P.A.)
58	APFNF-AAF	Data Stream	<i>lv</i>	1-Month A2/P2/F2 NCP (% P.A.) minus 1-Month Aa Financial Commercial Paper (% P.A.)
59	TED	Data Stream, FRED	<i>lv</i>	3Month Tbill minus 3-Month London Interbank Offered Rate (%)
60	MAaa-10YTB	Data Stream	<i>lv</i>	Moody Seasoned Aaa Corporate Bond Yield (% P.A.) minus Y10-Tbond
61	MBaa-10YTB	Data Stream	<i>lv</i>	Moody Seasoned Baa Corporate Bond Yield (% P.A.) minus Y10-Tbond
62	MLA-10YTB	Data Stream, FRED	<i>lv</i>	Merrill Lynch Corporate Bonds: A Rated: Effective Yield (%) minus Y10-Tbond
63	MLAA-10YTB	Data Stream, FRED	<i>lv</i>	Merrill Lynch Corporate Bonds: Aa Rated: Effective Yield (%) minus Y10-Tbond
64	MLAAA-10YTB	Data Stream, FRED	<i>lv</i>	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	<i>lv</i>	6-month treasury bill market bid yield at constant maturity (%) minus Fed Funds
71	1YTB-FF	Data Stream	<i>lv</i>	1-year treasury bill yield at constant maturity (% P.A.) minus Fed Funds

Table E.2 (Cont'd)

No.	Short Name	Source	Tran	Description
72	10YTB-FF	Data Stream	<i>lv</i>	10-year treasury bond yield at constant maturity (% P.A.) minus Fed Funds
73	6MTB-3MTB	Data Stream	<i>lv</i>	6-month treasury bill yield at constant maturity (% P.A.) minus 3M-Tbills
74	1YTB-3MTB	Data Stream	<i>lv</i>	1-year treasury bill yield at constant maturity (% P.A.) minus 3M-Tbills
75	10YTB-3MTB	Data Stream	<i>lv</i>	10-year treasury bond yield at constant maturity (% P.A.) minus 3M-Tbills
76	BKEVEN05	FRB	<i>lv</i>	US Inflation compensation: continuously compounded zero-coupon yield: 5-year (%)
77	BKEVEN10	FRB	<i>lv</i>	US Inflation compensation: continuously compounded zero-coupon yield: 10-year (%)
78	BKEVEN1F4	FRB	<i>lv</i>	BKEVEN1F4
79	BKEVEN1F9	FRB	<i>lv</i>	BKEVEN1F9
80	BKEVEN5F5	FRB	<i>lv</i>	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,

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

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 k th row and v th 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 $K \ll V$ (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,

$$\underbrace{\overbrace{w_{i,t+s}}^{V \times 1}}_{\text{word dist of article } i \text{ at time } t+s} \sim \text{Mult} \left(\underbrace{\overbrace{\hat{\Phi}'}^{V \times K}}_{\text{topic-word dist.}}, \underbrace{\overbrace{\hat{\theta}_{i,t+s}}^{K \times 1}}_{\text{topic dist.}}, \underbrace{N_{i,t+s}}_{\# \text{ of words}} \right). \quad (\text{E.2})$$

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

⁴In theory, maximum-likelihood estimation is possible but it is computationally challenging.

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 estimation 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 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 further 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,

$$F_{k,t} = \frac{\sum_i \hat{\theta}_{i,k,t}}{\# \text{ of articles at } t} \quad (\text{E.3})$$

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.

F Machine Learning

Machine Algorithm Details

The basic dynamic algorithm follows the five-step approach of Bianchi et al. (2022): 1. Sample partitioning, 2. Training, 3. Hyperparameter tuning and re-optimization, 4. Out-of-sample prediction, and 5. 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.

⁵The authors used Bayesian criteria to find 180 to be an optimal number of topics.

At time t , a prior sample of size $\dot{T}$ is partitioned into two subsample 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

$$y_{j,t+h} = G^e(\mathcal{X}^t, \beta_{j,h,t}^e) + \epsilon_{j,t+h}.$$

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 $1, \dots, t$, and $G^e(\cdot)$ is a machine learning estimator that can be represented by a (potentially) high dimensional set of finite-valued parameters $\beta_{j,h,t}^e$. Our main estimator for $G^e(\cdot)$ is the Long Short-Term Memory (LSTM) recurrent neural network $G^{\text{LSTM}}(\mathcal{X}^t, \beta_{j,h}^{\text{LSTM}})$. In the validation step, we also ask the algorithm to compare the LSTM forecast to one using a simpler Elastic Net (EN) estimator, $G^{\text{EN}}(\mathcal{X}^t, \beta_{j,h}^{\text{EN}})$. $\mathcal{X}_t$ 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 $\beta_{j,h,t}^e$ are estimated by minimizing the mean-square loss function over the training sample with L_1 and L_2 penalties

$$L(\beta_{j,h,t}^e, \mathcal{X}^{T_E}, \lambda_t^e) \equiv \underbrace{\frac{1}{T_E} \sum_{\tau=1}^{T_E} (y_{j,\tau+h} - G^e(\mathcal{X}^\tau, \beta_{j,h,t}^e))^2}_{\text{Mean Square Error}} + \underbrace{\lambda_{1,t}^e \sum_{k=1}^K |\beta_{j,h,t,k}^e|}_{L_1 \text{ Penalty}} + \underbrace{\lambda_{2,t}^e \sum_{k=1}^K (\beta_{j,h,t,k}^e)^2}_{L_2 \text{ Penalty}}$$

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 $\beta_{j,h,t}^e$ is a function of the data $\mathcal{X}^{T_E}$ and a non-negative regularization parameter vector $\lambda_t^e = (\lambda_{1,t}^e, \lambda_{2,t}^e, \lambda_{0,t}^{\text{LSTM}})'$ where $\lambda_{0,t}^{\text{LSTM}}$ 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 λ_t^e are estimated by minimizing the mean-square loss over pseudo-out-of-sample forecast errors generated from rolling regressions through the validation sample:

$$\begin{aligned}\hat{\lambda}_t^{EN}, \hat{T}_E, \hat{T}_V &= \underset{\lambda_t^{EN}, T_E, T_V}{\operatorname{argmin}} \left\{ \frac{1}{T_V - h} \sum_{\tau=T_E}^{T_E+T_V-h} \left(y_{j,\tau+h} - G^{EN}(\mathcal{X}_\tau, \hat{\beta}_{j,h,\tau}^{EN}(\mathcal{X}^\tau, \lambda_t^{EN})) \right)^2 \right\} \\ \hat{\lambda}_t^{LSTM}, \hat{T}_E, \hat{T}_V &= \underset{\lambda_t^{LSTM}, T_E, T_V}{\operatorname{argmin}} \left\{ \frac{1}{T_V - h} \sum_{\tau=T_E}^{T_E+T_V-h} \left(y_{j,\tau+h} - G^{LSTM}(\mathcal{X}_\tau, \hat{\beta}_{j,h,\tau}^{LSTM}(\mathcal{X}^\tau, \lambda_t^{LSTM})) \right)^2 \right\}\end{aligned}$$

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

The estimation of $\hat{\beta}_{j,h,t}^e(\mathcal{X}^t, \hat{\lambda}_t^e)$ is repeated sequentially in rolling subsamples, with parameters estimated from information known at time t . Note that the time t subscripts of $\hat{\beta}_{j,h,t}^e$ and $\hat{\lambda}_t^e$ 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 $\mathbb{E}_t^e[y_{j,t+h}]$, defined by

$$\mathbb{E}_t^e[y_{j,t+h}] \equiv G^e\left(\mathcal{X}^t, \hat{\beta}_{j,h,t}^e(\mathcal{X}^t, \hat{\lambda}_t^e)\right)$$

Finally, the machine MSE is computed by averaging across the sequence of squared forecast errors in the true out-of-sample forecasts for periods $t = (\dot{T} + h), \dots, T$ where T is the last period of our sample. The true out-of-sample forecasts used for neither estimation nor tuning is the *testing subsample* 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 $\mathbb{E}_t^e[y_{j,t+h}]$ 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 $\mathbb{F}_t[y_{j,t+h}]$ 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:

$$y_{j,t+h} = G_{jh}^e(\mathcal{X}^t) + \gamma_{jh\mathbb{M}}\mathbb{F}_t[y_{j,t+h}] + \epsilon_{j,t+h}, \quad h \geq 1 \quad (\text{F.1})$$

where $\gamma_{jh\mathbb{M}}$ is a parameter to be estimated, and where $G_{jh}^e(\mathcal{X}^t)$ represents a ML estimator as function of big data. Note that the intercept α_{jh} from Bianchi et al. (2022) gets absorbed into the $G_{jh}^e(\mathcal{X}^t)$ 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:

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

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

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

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 an 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:

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

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

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

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

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

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

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

where $n = 1, \dots, N$ 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 $\tilde{c}_t^n$. 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. $\sigma : \mathbb{R} \rightarrow \mathbb{R}$ is the sigmoid function: $\sigma(x) = (1 + e^{-x})^{-1}$. $\tanh : \mathbb{R} \rightarrow \mathbb{R}$ is the hyperbolic tangent function: $\tanh(x) = \frac{e^{2x}-1}{e^{2x}+1}$. The $\odot$ operator refers to elementwise multiplication. $\beta_{j,h}^{\text{LSTM}} \equiv (((\text{vec}(W^{(g^n h^{n-1})})', \text{vec}(W^{(g^n h^n)})', b'_{g^n})_{g \in \{c, f, i, o\}})_{n=1}^N, \text{vec}(W^{(y h^N)})', b_y)'$ 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 $\beta_{j,h}$ 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 $\beta_{j,h}^{(0)}$ randomly, where the superscript (0) in parentheses indexes the iteration for an estimate of $\beta_{j,h}^{\text{LSTM}}$.

- (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:

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

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

- (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 $t = 1, \dots, T_E$ 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 $1, \dots, M$
- 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

⁶When N_S/N_B is not a whole number, $\lceil N_S/N_B \rceil$ 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.

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:

$$\begin{aligned} \underbrace{\bar{h}_t^{n-1}}_{D_{h^{n-1}} \times 1} &= \underbrace{r_{t,h_t^{n-1}}^{(k)}}_{D_{h^{n-1}} \times 1} \odot \underbrace{h_t^{n-1}}_{D_{h^{n-1}} \times 1}, \quad r_{t,h_t^{n-1},i}^{(k)} \stackrel{iid}{\sim} \text{Bernoulli}(p_{h_t^{n-1}}), \quad i = 1, \dots, D_{h^{n-1}} \\ \underbrace{\bar{h}_{t-1}^n}_{D_{h^n} \times 1} &= \underbrace{r_{t,h_{t-1}^n}^{(k)}}_{D_{h^n} \times 1} \odot \underbrace{h_{t-1}^n}_{D_{h^n} \times 1}, \quad r_{t,h_{t-1}^n,i}^{(k)} \stackrel{iid}{\sim} \text{Bernoulli}(p_{h_{t-1}^n}), \quad i = 1, \dots, D_{h^n} \end{aligned}$$

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 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

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

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

(c) *Learning rate shrinkage.* Update the parameters to $\beta_{j,h}^{(k)}$ 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.

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

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

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

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

π^k denotes the k -th 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 Criteria.* Stop iterating and return $\beta_{j,h}^{(k)}$ if one of the following holds:

- *Early stopping.* At each iteration, use the updated $\beta_{j,h}^{(k)}$ 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 $\beta_{j,h}$ towards the initial guess $\beta_{j,h}^{(0)}$, 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 $\hat{\beta}_{j,h,T_E}^{LSTM}(\mathcal{X}^t, \boldsymbol{\lambda}^{LSTM}, R)$. 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 $\beta_{j,h}^{(0)}$, the sequences selected in each mini-batch $B^{(k)}$, and the dropout mask $r_t^{(k)}$.

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 to 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 hyperparameters 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 $D_{h^1}, \dots, D_{h^N}$ in each hidden layer are hyperparameters 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. We also select the L_1 and L_2 parameters iteratively in the cross-validation step.

Table F.1 reports the hyperparameters for the LSTM network and its estimation. Hyperparameters reported as a range or a set of values are cross-validated. The hyperparameters 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 F.1. 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 180 input variables, the candidate architectures range from 2,960 parameters (single-layer, 4 neurons) to about 9.8 million parameters (5-layer, 512 neurons per layer). For earnings growth prediction with 231 input variables, the architectures span from 22,208 parameters (4-layer, 16 neurons per layer) to about 9.9 million parameters (5-layer, 512 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

Table F.1: Candidate hyperparameters for the machine learning forecast

Variable	Earnings Growth	Earnings Growth	Earnings Growth	Stock Returns	Price Growth
Horizon	1-Year	4-5 LTG	1-10 LTG	1-Year	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, 128, 256, 512	16, 32, 64	16, 32, 64	16, 32, 64, 128, 256, 512	16, 32, 64, 128, 256, 512
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, ..., 20	1, ..., 20	1, ..., 20	1, ..., 20	1, ..., 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.

asking about $r_{t,t+12}^d$, 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 $r_{t,t+12}^d$, 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.
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 2023: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 and 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. (2026)). 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 $\mathbf{Z}_{jt} \equiv (y_{j,t}, \hat{\mathbf{G}}'_t, \mathbf{W}'_{jt})'$ is an $r = 1 + r_G + r_W$ vector which collects the data at time t with $\mathbf{Z}_{jt} \equiv (y_{j,t}, \dots, y_{j,t-p_y}, \hat{\mathbf{G}}'_t, \dots, \hat{\mathbf{G}}'_{t-p_G}, \mathbf{W}'_{jt}, \dots, \mathbf{W}'_{jt-p_W})'$ 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 93 real-time macro series; includes both monthly and quarterly series, with monthly series converted to quarterly according to the method described in the data appendix.
2. $\mathbf{G}_{F,t-k}$, for $k = 0, 1$ are factors formed from a financial data set $\mathcal{D}^F$ with 147 monthly financial series.
3. $\mathbf{G}_{D,t-k}^Q$, 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 $\mathbf{G}_{D,t}^Q(\mathbf{w})$ using a weighted average of daily factors, with the weights $\mathbf{w}$ dependent on two free

parameters that are chosen to minimize the sum of squared residuals in a regression of $y_{j,t+h}$ on $\mathbf{G}_{D,t}(\mathbf{w})$.

4. *LDA topics* $F_{k,t-j}$, for topic $k = 1, 2, \dots, 50$ 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, 3, 4, 6, 12, and 24 month-ahead federal funds futures (FFF) contract rate and the changes in the 1, 2, 3, 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 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 13 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 93 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 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 93 real-time macro series; includes both monthly and quarterly series, with monthly series converted to quarterly according to the method described in the data appendix.
2. $\mathbf{G}_{F,t-k}$, for $k = 0, 1$ are factors formed from a financial data set $\mathcal{D}^F$ with 147 monthly financial series.
3. $\mathbf{G}_{D,t-k}^Q$, 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 $\mathbf{G}_{D,t}^Q(\mathbf{w})$ using a weighted average of daily factors, with the weights $\mathbf{w}$ dependent on two free parameters that are chosen to minimize the sum of squared residuals in a regression of $y_{j,t}$ on $\mathbf{G}_{D,t}(\mathbf{w})$.
4. *LDA factors* $F_{k,t-j}$, for topic $k = 1, 2, \dots, 50$ 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, 3, 4, 6, 12, and 24 month-ahead federal funds futures (FFF) contract rate and the changes in the 1, 2, 3, 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,

⁷A 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

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 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 13 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 F.2: Number of Right-Hand Side Variables

	Stock Return	Earnings
Macro Factors	20	20
Financial Factors	20	20
Daily Factors	10	10
LDA Media-Sentiment Factors	100	150
FOMC Surprises	13	13
Macro Data Surprises	6	6
Other Predictors	11	12
Total	180	231

Notes: This table shows the number of right-hand-side input variables (including lags) for each forecast target, equal to the models' integrated-gradient input dimension.

G 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 $1, \dots, t$. The LSTM function estimated at each t can be arbitrarily well approximated as the sum of three components:

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

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 F, 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 $\hat{\beta}_t$, 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 by 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 i th 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

$$\nabla G(\cdot)_i \equiv \int_{\delta=0}^{\delta=1} \frac{\partial G(\delta \mathcal{X}_t; \cdot)}{\partial \mathcal{X}_{i,t}} d\delta \approx \frac{1}{M} \sum_{m=0}^M \frac{\partial G(\frac{m}{M} \mathcal{X}_t; \cdot)}{\partial \mathcal{X}_{i,t}}, \quad i = 1, \dots, D. \quad (\text{G.2})$$

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 G.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 LSTM estimator as a function of 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:

$$\hat{\lambda} = \underset{\lambda}{\operatorname{argmin}} \left\{ \frac{1}{4} \sum_{h=1}^4 \left[\frac{1}{T_V} \sum_{\tau=1}^{T_V} (\mathbb{F}_{\tau+h}[y_{j,\tau+h+4}] - G(\cdot))^2 \right] \right\} \quad (\text{G.3})$$

$$+ \underbrace{\lambda_1 \sum_{h=1}^4 \sum_{k=1}^K |\hat{\beta}_{\tau,k}|}_{L_1 \text{ Penalty}} + \underbrace{\lambda_2 \sum_{h=1}^4 \sum_{k=1}^K (\hat{\beta}_{\tau,k})^2}_{L_2 \text{ Penalty}} \quad (\text{G.4})$$

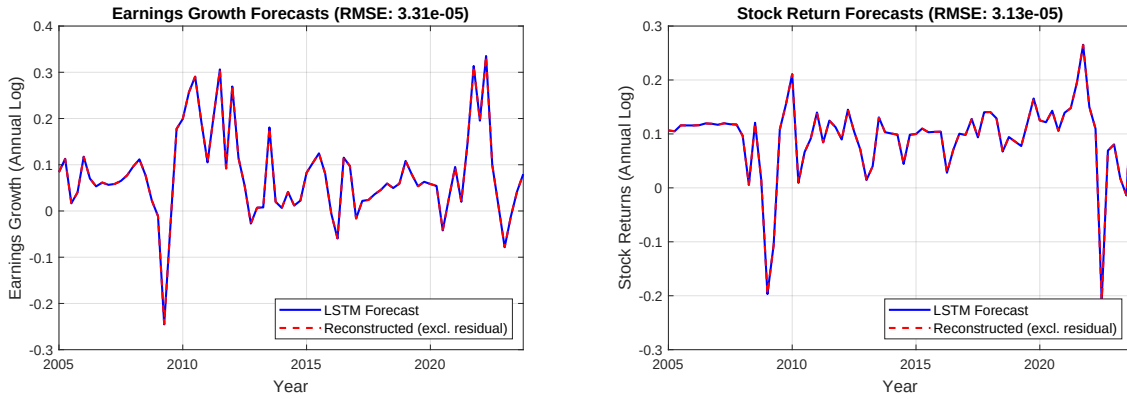
In the loss function, $h \in \{1, 2, 3, 4\}$ indexes the forecast horizons for the predictive model. $\mathbb{F}_{\tau+h}[y_{j,\tau+h+4}]$ denotes the survey forecast made at time $\tau + h$ for the value of variable y_j four quarters ahead, i.e., for period $\tau + h + 4$. In other words, at date $\tau + h$, 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 $\hat{\beta}_{\tau}$ are the corresponding network parameters estimated using training data and shared regularization penalties λ_1 and λ_2 . All other LSTM-related hyperparameters 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 survey forecast takes the form below, where the first three terms are the final fitted survey model obtained by the machine algorithm:

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

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

Figure G.1: Accuracy of Forecast Decomposition



Notes: This figure validates the accuracy of the integrated gradients decomposition by comparing the original LSTM forecasts (blue solid line) with the forecasts reconstructed from the interpretable decomposition components only, namely the local mean, nonlinear history, and current-information themes, excluding the residual (red dashed line). The Root Mean Square Error (RMSE) reported in the title is the discrepancy between the two, which equals the residual left uncaptured by the integrated gradients decomposition; it is negligible relative to the variation in the forecasts (RMSE of 3.1e-05 for returns and 3.3e-05 for earnings, against forecast standard deviations of 0.072 and 0.094). The x-axis indicates the forecasted period with realized outcomes. Stock returns are expressed as log annual returns and earnings growth as the log annual growth rate. The sample is the testing sample, 2005:Q1-2023:Q4. The training-sample fits are excluded: in sample the reconstruction is exact to machine precision by construction, so pooling those quarters would understate the approximation error rather than measure it.

H 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 output or unemployment. Following Ludvigson and Ng (2007), we relate the factors to the

Table H.1: 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

Notes: [†]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.

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 H.2 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 k th 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 k th topic is a V -dimensional vector $\phi_{k,t}$ in the k th row of the topic-word distribution $\Phi_t = [\phi_{1,t}, \dots, \phi_{K,t}]'$, 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 probability across our testing subsample 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 a testing subsample of 2005:Q1 to 2023:Q4. Therefore, 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 H.2: 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

Notes: 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.

Table H.3: 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.