New News is Bad News

Paul Glasserman* Harry Mamaysky† Jimmy Qin‡

August 29, 2023

Abstract

An increase in the novelty of news predicts negative stock market returns and negative macroeconomic outcomes over the next year. We quantify news novelty – changes in the distribution of news text – through an entropy measure, calculated using a recurrent neural network applied to a large news corpus. Entropy is a better out-of-sample predictor of market returns than a collection of standard measures. Cross-sectional entropy exposure carries a negative risk premium, suggesting that assets that positively covary with entropy hedge the aggregate risk associated with shifting news language. Entropy risk cannot be explained by existing long-short factors.

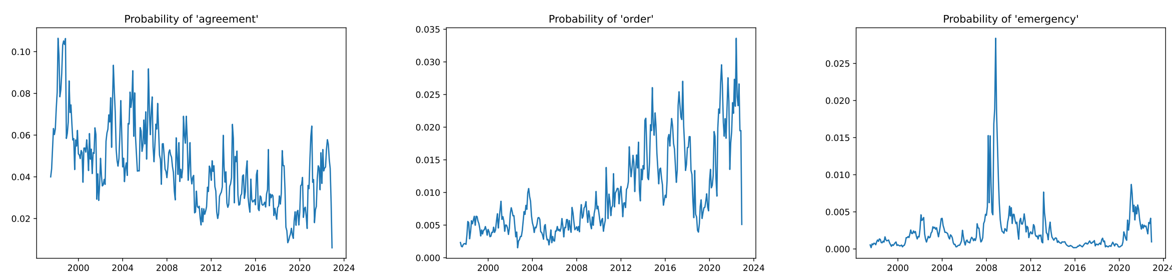
Keywords: entropy; natural language processing; news articles; empirical asset pricing
JEL Codes: E30, G12, G14, G17

*Columbia Business School, pg20@columbia.edu.
†Columbia Business School, hm2646@columbia.edu.
‡Columbia Business School, qq2127@columbia.edu. Supported by the W. Edwards Deming Center for Quality, Productivity and Competitiveness at Columbia Business School.

1 Introduction

Several studies have documented that the sentiment of news text forecasts short-term changes in asset prices, with negative news forecasting negative returns. We find that a change in the distribution of news text — a measure of the novelty or unusualness of the news rather than its sentiment — forecasts negative market returns and macroeconomic outcomes over the subsequent year. Consistent with this pattern, we find that assets that positively covary with our measure carry a negative risk premium: investors accept lower compensation to hold assets that hedge the risk associated with a shift in the language of news.

To motivate the idea, consider the string of words "the securities and exchange commission issued an." What is the likelihood that the next word in the sentence is "agreement," "order," "emergency," or any other English language word? Using a recurrent neural network re-estimated in rolling windows over 1.6 million Reuters news articles spanning a period of 27 years, we can quantify exactly how this distribution evolves over time. For example, the likelihood that the next word is "agreement" has a decreasing time trend, while the likelihood that the next word is "order" has an increasing time trend. The likelihood that the subsequent word is "emergency" spikes during the financial crisis and again during the COVID-19 pandemic.[1] Although in principle such shifts in language structure could be neutral, we find that they generally contain negative news about markets and the economy. This finding leads us to explore and subsequently uncover a rich structure of risk pricing associated with exposures to the changing distribution of news language.



The probability of "agreement", "order", and "emergency" following the string of words "the securities and exchange commission issued an." The distributional shift over time is estimated by a recurrent neural network.

We refer to our measure as entropy; it can also be seen as a measure of the cross entropy (or dissimilarity) between the current and past distributions of news text. Glasserman and Mamaysky (2019) used an entropy measure to forecast volatility, rather than the

[1] The methodology to calculate these conditional probabilities is explained in Section 3.4.

direction of returns. To estimate conditional word probabilities, they used the empirical frequencies of word sequences in what is known as an n-gram method in the natural language processing (NLP) literature; see Jurafsky and Martin (2023). Exact phrases, like in the above example, occur rarely if they are more than a few words long, limiting the accuracy of the n-gram method.

Here we use a far more powerful recurrent neural network (RNN) model with a long short-term memory (LSTM) architecture (Mikolov et al. 2010, Hochreiter and Schmidhuber 1997) to estimate conditional word probabilities and thus to calculate entropy. Whereas in practice the n-gram approach conditions on only three or four words prior to the predicted word, an RNN model captures far more contextual information in assigning probabilities to words. We apply the RNN model with word embeddings that encode rich information about the relationships between words, adding to the power of the method. Unlike the latest and most complex large language models (like GPT-4, BERT, and their relatives), our RNN can be trained relatively quickly. This allows us to retrain our RNN monthly in rolling windows over our news data, and then use the most recent model to assign an entropy score to the next month's news articles, which reflects the degree to which the text in these articles deviates from the RNN model's conditional word distributions. We can then assess the in- and out-of-sample forecasting performance of our entropy measure.

Each month we calculate ENT, the change in our raw entropy measure over the previous 12 months. For simplicity, we often refer to ENT simply as *entropy* in the rest of the paper. In in-sample regressions, we find that a one standard deviation increase in ENT forecasts a 3% decline in cumulative S&P 500 returns over the next 12 months, even after controlling for a large number of alternative predictors. The effect is highly statistically significant.

To check out-of-sample predictability, we compare univariate forecasts for a large number of predictors; the out-of-sample R-squared for ENT is positive at multiple horizons but overwhelmingly negative for the other predictors. In particular, entropy is the best out-of-sample predictor of market returns when compared to a long list of variables proposed in prior studies, including the inverse of the cyclically adjusted price-to-earnings ratio of Campbell and Shiller (1988) $CAPE$, the variables tested in Welch and Goyal (2008), Campbell and Thompson (2008), and David and Veronesi (2022), as well as economic policy uncertainty (EPU, Baker et al. 2016), the VIX (Martin 2017), and the aggregate consumption to wealth ratio of Lettau and Ludvigson (2001). We are careful to estimate our RNN model in rolling windows to ensure the test is truly out-of-sample. Given the extensive literature devoted to identifying successful out-of-sample risk premium forecasters, the out-of-sample forecasting performance of ENT is remarkable.

If an increase in ENT forecasts a decline in the stock market, then an asset that positively covaries with ENT provides a hedge against this deterioration in the investment opportunity set. We, therefore, expect ENT to carry a negative risk premium: investors will accept lower expected returns on assets that covary positively with ENT because of the hedge they provide. This ICAPM-style (Merton 1973) argument provides the intuition underlying the sign consistency property proposed by Maio and Santa-Clara (2012). Through cross-sectional regressions that control for standard factors, we estimate a monthly risk premium for ENT that is indeed negative, ranging from -0.06% to -0.09% per unit standard deviation of cross-sectional entropy loading, depending on the choice of test assets. While the magnitude of this risk premium is in line with that of the other factors, none of the other cross-sectional pricing factors satisfy the Maio and Santa-Clara (2012) consistency property that a factor's market return forecasting direction and cross-sectional risk premium have the same sign.

The structure of our ENT measure allows its decomposition into two components, which we refer to as *news updates* and *model updates*. The first component measures the change in entropy, holding the distributional model of text fixed, and the second measures the change in the distributional model. The first component is thus more focused on the novelty of the text in the current month. We find that the predictive power of entropy stems primarily from the first component, which negatively forecasts future returns out-of-sample and carries a negative cross-sectional risk premium.

Entropy and its news innovation component forecast year-ahead macroeconomic outcomes. Increased entropy is associated with future increases in unemployment and the VIX, and with future declines in industrial production, inflation, interest rates, and corporate earnings. Consistent with an ICAPM-style hedging argument, market participants are willing to accept lower expected returns for securities that hedge entropy risk. However, entropy also negatively forecasts aggregate market returns suggesting that markets do not fully reflect the information content of entropy either because of investor informational constraints as in Sims (2003) or because of slow-moving institutional capital and limits to arbitrage (Gabaix and Koijen 2021, Gromb and Vayanos 2010). While improving data and analytics may alleviate the former channel, the latter institutional-constraint channel is likely to persist.

Lastly, we show that the replicating portfolio for our entropy measure cannot be explained by the multitude of existing long-short factors (Harvey et al. 2016, Feng et al.

2020). Jensen et al. (2021) argue that most factor-based asset pricing results can be replicated, and in doing so construct a database of over 150 long-short factor portfolios.[2] Using this database of over 150 long-short factors we show that our entropy-change replicating portfolio is among the long-short portfolios that are most poorly explained by existing factors. Furthermore, entropy appears largely unrelated to several measures of economic uncertainty (Baker et al. 2016, Bekaert et al. 2022, Jurado et al. 2015, Azzimonti 2018, David and Veronesi 2022) and contains information that is distinct from news sentiment. Entropy, with its sign-consistent price of risk and out-of-sample forecasting performance, is also largely unspanned by existing long-short factors and measures of uncertainty.

1.1 Literature Review

Our paper is related to the literature examining the impact of text similarity in financial markets. Tetlock (2011) measures news staleness as the similarity of news stories to prior news stories about the same firm, and finds that firms' stock returns are less responsive to stale news but that retail investors overreact to stale news. Cohen et al. (2020) show that firms with quarter-over-quarter changes in the language of their 10-Ks and 10-Qs experience significantly lower future stock return relative to non-changers, which they interpret as market underreaction to the information content of corporate reports. Hoberg and Phillips (2016, 2018) construct firm peer groups using textual similarity of product descriptions contained in 10-K reports and show that there are cross-momentum effects within text-based peer groups.

The traditional definition of dissimilarity has focused on differences in word counts between small groups of related documents. Glasserman and Mamaysky (2019) extend this concept to measure how unusual an article's language is relative to the text of *all* prior articles. They show that, at the firm-level and in aggregate, unusual news forecasts increases in future realized and implied volatility. Prior measures of dissimilarity or informativeness suffer from several problems: word counts cannot account for article context or synonyms, while the n-gram approach of Glasserman and Mamaysky (2019) suffers from the sparsity issue that many feasible n-grams are never observed, especially for large n. Our entropy methodology offers a dramatic improvement over prior efforts by using word embeddings inside a deep-learning framework.

Our paper is related to a long literature in computer science and linguistics which aims to create probabilistic models for text using deep learning. Hochreiter and Schmidhuber

[2] Chen and Zimmerman (2021) make a similar point and also provide factor-replication code and data.

(1997) introduce the long short-term memory (LSTM) architecture, which revolutionized the ability of recurrent neural networks to successfully model large text corpora (Charniak 2019). Pennington et al. (2014) introduces the GloVE model of word embeddings which map words into a high-dimensional space where vector operations reflect the semantic content of words. While the transformer underlying ChatGPT (Radford et al. 2019) is a far larger model than the one we use, it has been trained on an extensive corpus obtained from the internet, and therefore cannot be used for backtesting trading strategies because the model contains future information when applied to historical data. Our approach, while restricted to a smaller model and data set, is trained in rolling windows, and as such, it makes forecasts using only data that would have been available to market participants in real-time, which can be used for out-of-sample testing.

Several papers are concerned with aggregate measures of text or news flow. Jiang et al. (2019) show that an index of aggregate manager sentiment, based on the text of firms' 10-Ks, 10-Qs, and conference calls, negatively forecasts stock returns, suggesting stock investors overreact to the information content of management communication. Using a dictionary-based approach, Shapiro et al. (2022) construct an aggregate economic sentiment measure from articles in 16 major U.S. newspapers and show that positive economic news is associated with growth in future consumption, output, and real rates. Baker et al. (2016) introduce the economic policy uncertainty (EPU) measures which count the number of times terms like *uncertainty*, *economic*, and *legislation* (and other similar word combinations) appear in close proximity to each other in articles from 10 major U.S. newspapers. They then use a weighted average of the frequency of major news discussing economic policy uncertainty, expiring tax provisions, and forecaster disagreement about government purchases to construct their measure. They show that EPU negatively forecasts investment, output, and employment, and at the firm-level EPU predicts "greater stock price volatility and reduced investment and employment" for firms most heavily exposed to government policy.

Brogaard and Detzel (2015) test the asset pricing implications of EPU and show that it positively forecasts market returns but carries a negative cross-sectional price of risk.[3] Based on these findings, EPU does not satisfy the Maio and Santa-Clara (2012) sign property which suggests that if EPU positively forecasts future returns, it should have a positive cross-sectional price of risk. When controlling for other forecasting variables and

[3]Lin (2022) is a related paper that shows that another economic uncertainty index positively forecasts stock market volatility, but does not earn a significant price of covariance risk in a cross-section of 25 Fama-French size and book-to-market sorted portfolios, which is also a violation of the ICAPM sign property proposed in Maio and Santa-Clara (2012).

for *ENT*, we also find that *EPU* forecasts returns positively and that it carries a negative risk premium, but the forecasting results are not consistent and don't hold out-of-sample, and the negative risk premium is not significant in all specifications. In contrast, *ENT* is a significant return forecaster in- and out-of-sample, satisfies the ICAPM-style sign property, and has a significant price of risk across different specifications.

It is natural to contrast our entropy measure with measures of model uncertainty. Among others, Hansen and Sargent (2008), Aït-Sahalia et al. (2021), and Brenner and Izhakian (2018) link risk premia to uncertainty, distinguishing uncertainty about the true model and ambiguity about the true outcome probabilities from the market or economic volatility. An increase in *ENT* might suggest heightened uncertainty, but theory commonly associates uncertainty with a positive risk premium, and we find a negative risk premium for *ENT*.

The remainder of this paper proceeds as follows. Section 2 discusses the data we use. In Section 3, we define the entropy measure and compare several estimation methods. Section 4 provides an empirical assessment of the forecasting power of entropy for future market returns. Section 5 investigates the cross-sectional association between aggregate entropy and returns. Section 6 decomposes entropy into news and model innovation components, investigates the macroeconomic forecasting properties of entropy, and discusses potential channels for why entropy forecasts market returns. Section 7 offers some robustness checks and Section 8 concludes. Technical details and supplementary results are included in an Online Appendix.

2 Data and Variable Construction

We construct our entropy measure using the Thomson Reuters News Feed Direct archive from January 1996 to December 2022, which is the time frame of our analysis. We removed articles that represent multiple rewrites of the same initial story, retaining only the first article in a given chain.[4] We only kept news articles that were written in English and discussed S&P500 companies.[5] We excluded articles with headlines containing certain keywords such as "shh margin trading", "nyse", "imbalance", "machine generated", and "research alert;" these keywords typically signal messages that are not really news articles

[4]Thomson Reuters tracks articles by assigning a Primary News Access Code (PNAC). Articles that share the same PNAC are duplicates. This generally happens when there is an update to the coverage of the same event.

[5]A detailed description of the article selection methodology is in Glasserman et al. (2023). The methodology involves mapping Reuters company names to CRSP company names using fuzzy matching.

or exhibit irregularities. For example, "machine generated" articles were produced using algorithms or automated processes, rather than being created manually by human writers and editors. As such, their content may not have the same level of context, nuance, or analysis as content created by human writers. "[R]esearch alert" articles present hundreds of duplications. We further discarded articles with fewer than 30 words. This filtering process leaves us 1,642,517 articles. Figure 1 depicts the number of articles and average article length per month, respectively. The monthly article count increased from 1996 and peaked in the early 2000s, remaining steady until 2010 before gradually decreasing. The average number of words per article steadily increased from 1996, reaching an approximate steady-state level of 250 words per article in the early 2000s, with fluctuations observed after 2013.

We retrieve the Chicago Board Options Exchange's CBOE Volatility Index (*VIX*) from the Federal Reserve Economic Data (FRED) server. We write *VIX2* for the square of *VIX*. We include additional uncertainty measures from Bekaert et al. (2022) (*BEX*) and Jurado et al. (2015) (*JLN*), and the Azzimonti (2018) Partisan Conflict Index (*PCI*), all of which can be retrieved from the David and Veronesi (2022) review article data set. As a proxy for the risk-free rate, we use the market yield on U.S. Treasury securities at ten-year (*DGS10*) and two-year (*DGS2*) constant maturity, also obtained from FRED. We construct the 2s-10s spread, *DGS10-2*, as *DGS10* minus *DGS2*. As conditioning variables for stock returns, we include the monthly S&P 500 dividend yield (*DY*), monthly cyclically adjusted PE ratio for the S&P 500 (*CAPE*), and the consumption-wealth ratio (*CAY*). We also use daily and monthly returns of Fama and French (2015) 5 factor model (*MKT*, *SMB*, *HML*, *RMW*, *CMA*) and the momentum factor (*UMD*).[6] *MKT* refers to the CRSP value-weighted index return net of the risk-free rate. We write *Return1* for the market return of the previous month, *Return12* for the cumulative market return of the previous 12 months excluding the most recent month, and *Return60* for the cumulative market return of the previous 60 months including the most recent month. Similarly, we construct *SMB60* (*HML60*, *RMW60*, *CMA60*, *UMD60*) as the *SMB* (*HML*, *RMW*, *CMA*, *UMD*) factor return of the previous 60 months. We use the last 60-month lagged factor returns in the forecasting regressions in Section 4 to convert factor returns to state variables as suggested by Maio and Santa-Clara (2012).[7]

[6]They are obtained from the Nasdaq Data Link `https://data.nasdaq.com/`, Robert Shiller's website `http://www.econ.yale.edu/~shiller/data.htm`, Martin Lettau's website `https://sites.google.com/view/martinlettau/data`, and Ken French's website `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/biography.html` respectively.

[7]Maio and Santa-Clara (2012) use 60-month lagged returns for two factors, but use a different construction for converting *SMB* and *HML* to state variables. For consistency, we use the 60-month lagged

3 The Entropy Measure

We build upon and extend the idea from Glasserman and Mamaysky (2019) to construct an entropy score for the unusualness of news text as a market signal. A text is considered unusual if it has a low likelihood relative to a model of language probability. This problem has been studied in the natural language processing literature on word prediction; see, in particular, Chapter 3 of Jurafsky and Martin (2023).

3.1 The Entropy of Text

Our goal is to estimate the probability of a new set of articles (an *evaluation text*) under a probability model $\mathbb{P}$ estimated from past articles (referred to as the *reference text* or *training corpus*). We can represent an evaluation text as a sequence of N words $w_1 w_2 \cdots w_N$. Its probability is given by the product of conditional probabilities

$$\mathbb{P}(w_1 \cdots w_N) = \prod_{k=1}^{N} \mathbb{P}(w_k | w_1 \cdots w_{k-1}), \tag{1}$$

in which the first factor is $\mathbb{P}(w_1)$, the unconditional marginal probability of word w_1. The average negative log probability per word is then given by

$$-\frac{1}{N} \ln \mathbb{P}(w_1 \cdots w_N) = -\frac{1}{N} \sum_{k=1}^{N} \ln \mathbb{P}(w_k | w_1 \cdots w_{k-1}). \tag{2}$$

If $\mathbb{P}$ correctly models the word-generating process, and if this process is stationary and ergodic, then (2) converges to the entropy of the process as N increases (see Section 3.8 of Jurafsky and Martin 2023). For fixed N, (2) measures the cross-entropy between the empirical distribution of the evaluation text and the model $\mathbb{P}$. For brevity, we will refer to (2) as the entropy of the evaluation text $w_1 \cdots w_N$. High entropy means low probability and thus signals an unusual text, relative to $\mathbb{P}$.

To calculate entropy, we need a model or estimate for $\mathbb{P}$. The n-gram approach (often with n equal to four or five) truncates the conditioning on the right side of (2) to the $n-1$ words immediately preceding w_k and thus misses additional context provided by words that came earlier. The n-gram approach estimates conditional probabilities of the form $\mathbb{P}(w_k | w_{k-n+1} \cdots w_{k-1})$ as the ratio of the number of occurrences of the sequences $w_{k-n+1} \cdots w_{k-1} w_k$ and $w_{k-n+1} \cdots w_{k-1}$, with adjustments to the numerator and denom-

returns for all factors.

inator for strings that were never previously observed, which we refer to as the *sparsity* issue. Choosing a larger n captures more conditioning information but makes it more likely that the conditioning sequence has rarely or never been observed before. The inability to condition on a large number of prior words leads to *approximation* errors in (2).

3.2 A Neural Network Approach

A neural network approach to modeling the probability $\mathbb{P}$ addresses the two shortcomings (approximation and sparsity) of the n-gram method. Recurrent Neural Networks (RNNs) are a type of neural network specifically designed for modeling sequential data, such as language. RNNs process input sequences one element at a time while maintaining an internal state that captures the context of the preceding elements in the sequence. This enables the network to capture complex dependencies between elements in the sequence, making them particularly useful for tasks like language modeling, where understanding the context of a word is critical for predicting the subsequent word in the sequence. We use an RNN with an LSTM (long short-term memory) architecture to incorporate additional conditioning information. Charniak (2019) provides an excellent introduction.

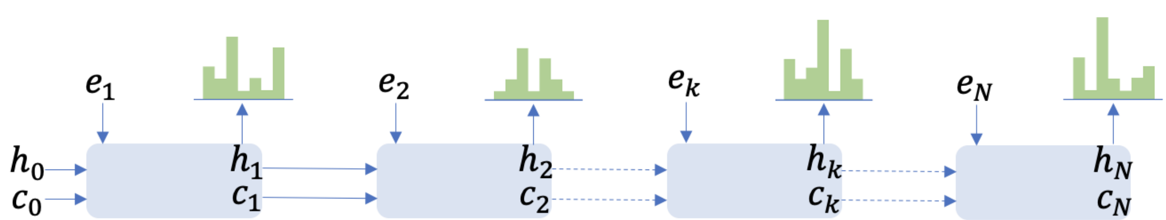


Illustration of RNN model with LSTM architecture. h contains the model's hidden state and c contains the long-term memory.

The above figure illustrates the evolution of the RNN model. The state of the RNN has two components, h and c. The h component is a vector of numerical values summarizing the context of the input sequence processed up to that point. It serves as a summary of the past inputs seen by the RNN up to the current time step and is updated iteratively as the RNN processes each word in the input sequence. The cell state, denoted by c, is an internal memory that stores and propagates information across time steps to capture long-term dependencies; this is the distinguishing feature of the LSTM structure. It allows the RNN to learn and remember long-term dependencies by regulating the flow of information through a mechanism called the "gates." The cell state is updated in parallel with the hidden state and is used to decide what information to retain, forget, or update.

The state of the RNN is updated after each word is read. Each word is represented through a word embedding (shown in the figure as the e), which is a vector of numerical values that encodes information about the meaning and use of the word. We use the well-known GloVe word embeddings (Pennington et al. 2014), which map each word to a 100-dimensional vector.[8]

At each step, the current state (h, c) determines a probability distribution over the model's vocabulary, which assigns a probability to each possible next word. The RNN thus embodies a model of $\mathbb{P}$. Once the RNN is trained, we evaluate the probabilities of the form $\mathbb{P}(w_k|w_1 \cdots w_{k-1})$ by feeding the embeddings of $w_1, \ldots, w_{k-1}$ into the RNN and then evaluating the RNN's conditional distribution $\mathbb{P}(\cdot|w_1 \cdots w_{k-1})$ at each observed word w_k. Unlike the n-gram approach, the conditional probabilities returned from the RNN reflect information from a very large number of prior words in the document.

Each of the arrows in the LSTM figure corresponds to a combination of linear and nonlinear transformations involving a large number of parameters. These are described in more detail in Section A2 of the Online Appendix. Training the RNN means calibrating these parameters to a reference or training text, a process we turn to next.

3.3 Model Training and Updating

Our neural network consists of an embedding (input) layer, an LSTM layer, and a fully connected (output) layer. The embedding layer maps words to their vector representations. The LSTM layer models temporal dependencies between words in a document. We set the dimension of the h and c vectors at 16. The LSTM layer, whose architecture is detailed in Section A2 of the Online Appendix, contains four W matrixes which transform word embeddings into an internal state and have dimension 16×100, four U matrixes which transform the internal state and have dimension 16×16, and four bias vectors of dimension 16 each. The total number of parameters in the LSTM layer is $4 \times (1600 + 256 + 16) = 7{,}488$. The output of the LSTM consists of a fully connected layer which maps the final hidden state h into a probability distribution over the vocabulary of 10,000 words. This layer contains a $10,000 \times 16$ matrix U_s which transforms the state h to 10,000 outputs and a 10,000-length bias vector for a total of 170,000 parameters. Thus the network has 177,488 parameters which must be trained.

All articles in the corpus are preprocessed by removing all punctuation strings and

[8] We obtain these from `https://nlp.stanford.edu/projects/glove/`.

converting text to lowercase. We retain only the 10,000 most frequent words in the whole corpus, representing all other words as "UNK" for unknown words. We chunk each article into segments of 100 words.[9] Training the RNN uses mini-batches, which are subsets of the entire dataset. The model's likelihood function is evaluated on the mini-batch, and parameter updates occur using the average gradient from each mini-batch item. This process requires specifying a batch size and the number of epochs. The batch size is the number of 100-word segments in a mini-batch. A larger batch size typically results in more stable updates to the weights but requires more memory and may lead to longer training times. Conversely, a smaller batch size allows for faster training times but may cause more fluctuations in weight updates. We choose a batch size of 128, i.e., the model updates its parameters after processing each mini-batch of 128 samples, each of length 100 words. An epoch is a full pass through the training set, during which the network processes every sequence of words in the training set once. Increasing the number of epochs can lead to better convergence of the network weights and improved performance on the validation set, but it also increases the risk of overfitting. We choose the number of epochs to be 50. These values for the batch size and number of epochs are standard (see Goodfellow et al. 2016, Keskar et al. 2016).

At the beginning of each epoch, the entire training dataset is randomly shuffled and divided into mini-batches without replacement.[10] During each epoch, the model processes one mini-batch at a time and updates its weights using backpropagation for each length-100 element of the mini-batch.[11] This process is repeated for all the mini-batches in the dataset until the entire dataset has been used for training in that epoch. Since the dataset is shuffled at the beginning of each epoch, the composition of the mini-batches changes from epoch to epoch, providing the model with a diverse set of samples to learn from. This shuffling means that, during training, the network state from 100-word sequence A that precedes 100-word sequence B in the same article will not carry over from A to B.

We initialize the training using the first six months of data (January 1996 – June

[9] For segments containing fewer than 100 words (at the end of articles, for example), we pad zero vectors to the actual words. When a batch of sequences is passed through the model, fixed-size length sequences allow the model to use vectorized operations and take advantage of hardware optimizations (i.e., parallel processing on GPUs). This results in faster training times and more efficient use of available computational resources.

[10] If the dataset is not perfectly divisible by the batch size, only the last batch in an epoch will contain fewer samples than the specified batch size.

[11] We train the model using the categorical cross-entropy loss function and the Adam optimizer (Kingma and Ba 2014). The optimization uses gradient descent with backpropagation: the loss function is computed for each step in the sequence of words in the training text, and gradients are backpropagated through the entire sequence, updating the weights of the model at each time step to reduce the loss function Werbos (1990).

1996) to compute the entropy for all articles in the following month (July 1996). Write $m_{[t-6,t-1]}$ for the RNN model trained using articles in the six months prior to t, and write $m_{[t-6,t-1]}(t)$ for the equal-weighted average entropy score of articles in month t using the RNN model trained over the prior six months. We do this to detect month t articles that may seem unusual based on information known only prior to month t.

Suppose in month t we have n_t articles in the archive, and the i-th article, represented by $w_{ti1}w_{ti2}\cdots w_{tiN_{ti}}$ of length N_{ti}. By forming the sample counterpart of (2), the entropy of this article calculated using model $m_{[t-6,t-1]}$ is

$$H_{ti} = -\frac{1}{N_{ti}}\sum_{k=1}^{N_{ti}} \ln(p_{tik}^{x_{tik}}) \tag{3}$$

where p_{tik} is the output of a vector of length 10,000 when we feed $w_{ti1}w_{ti2}\cdots w_{ti(k-1)}$ into $m_{[t-6,t-1]}$, x_{tik} is the position of word w_{tik} in the vector, and $p_{tik}^{x_{tik}}$ is the estimated probability of the k-th word being w_{tik} given the preceding words. Thus,

$$m_{[t-6,t-1]}(t) = \frac{1}{n_t}\sum_{j=1}^{n_t} H_{tj} = -\frac{1}{n_t}\sum_{j=1}^{n_t}\frac{1}{N_{tj}}\sum_{k=1}^{N_{tj}} \ln(p_{tjk}^{x_{tjk}}) \tag{4}$$

Unlike in model training, we keep track of the internal network state (h, c) in calculating all $p_{tjk} \in \mathbb{R}^{10,000}$ vectors when assigning probabilities to the words in article j in month t. The state of the RNN is then reset to zero before calculating the entropy of article $j + 1$.

From August 1996 onward, in order to calculate the entropies for all articles in month t, we first retrain the model $m_{[t-7,t-2]}$ to obtain an updated model $m_{[t-6,t-1]}$. The retraining starts from the parameter values obtained in $m_{[t-7,t-2]}$ and updates the values by running the parameter optimization on new text in month $t-1$ together with randomly sampled text from month $t-2$ to month $t-6$. In particular, we take all articles from month $t-1$, half of the articles randomly sampled from month $t-2$, one-quarter of the articles randomly sampled from month $t-3$, and so on until $\frac{1}{2^5}$ of the articles are randomly sampled from month $t-6$. Each retraining happens with batches of 128 length-100 word sequences with 50 epochs. Our main variable of interest, monthly entropy or ENT_t, is the difference between the average entropy in month t and month $t-12$, using the most recently available model at each time, or

$$ENT_t \equiv m_{[t-6,t-1]}(t) - m_{[t-18,t-13]}(t-12) \tag{5}$$

Our first observation is for July 1997.

The major benefit of using a relatively small language model, with "only" 177,488 parameters, is that we are able to estimate it in rolling windows, which means that ENT_t would have been available to market participants in real-time.[12] Furthermore, ENT_t can be decomposed into a new information component and a model update component, a feature of the model we exploit in Section 6.

3.4 Data Overview

Table 1 shows the summary statistics of variables used in subsequent analysis. All our analysis takes place at a monthly frequency. *ENT* ranges from -0.140 to 0.188, with a mean close to 0 and a standard deviation of 0.061. The table also shows statistics for the Economic Policy Uncertainty (*EPU*) index of Baker et al. (2016) and the San Francisco Fed's News Sentiment (*SEN*) index introduced by Shapiro et al. (2022). *EPU* is a well-known measure that is plausibly related to *ENT*, and thus we pay special attention to *EPU* in our analysis. *SEN* is also a related news-based measure. The construction of the *EPU* and *SEN* indexes is described in Online Appendix Section A3.

The time series of *ENT* and *EPU* are plotted in the top-left panel of Figure 2 as solid dark red and dashed light red lines, respectively. While *ENT* and *EPU* sometimes exhibit similar patterns (e.g., during the 2007–08 financial crisis), they often diverge. The time series of *ENT* and *SEN* are plotted in the top-right panel of Figure 2 as solid dark red and dashed light red lines, respectively. The peaks in *ENT* mostly coincide with either peaks or troughs in the sentiment measure. For example, during the 2007–08 financial crisis period, *SEN* is very negative while *ENT* is high. The time series of *ENT* and *VIX* are plotted in the bottom-left panel of Figure 2 as solid dark red and dashed light red lines, respectively. The peaks of *VIX* often coincide with peaks of *ENT*, particularly during the financial crisis period.

Figure 3 reports the average contemporaneous correlations between *ENT* and other control variables. *ENT* is not very correlated with any other control variables; its strongest correlation (0.394) is with *DY*. *ENT* has little correlation with *EPU* (0.115) and *SEN* (-0.280). The bottom-right panel of Figure 2 plots *ENT* and 12-month ahead cumulative returns. There appears to be some tendency for large positive next-12-month returns to occur at times of low entropy, and for low future returns to occur at times of moderate to high entropy. We investigate this tendency more formally in Section 4.

[12] Running on a GPU server with four NVIDIA T4 cards requires 21 hours to estimate the model in all rolling windows.

Figure 2 shows that January 2007, November 2010, and October 2014 were all months with large ENT peaks (and, interestingly, all of these were also low EPU months). Table 2 shows sample articles with high entropies from each of these months. We saw in the example of Section 1 how the conditional probabilities of words can change over time. We now give examples of how these changes contribute to the peaks in ENT through the entropy calculated in (3):

- In January 2007, there were several high-entropy news articles related to Fannie Mae and Freddie Mac. In one such article, the phrase "lack of progress in reining in mortgage lenders Fannie Mae and Freddie Mac leaves the economy at" was followed by the word "risk." But in January 2007, the word "risk" was perceived by the model to be unlikely in the context of this phrase, giving the sentence high entropy.[13] As illustrated in the middle panel of Figure 4, the conditional probability of "risk" increased dramatically through the mortgage crisis in late 2007 and the conservatorship of Fannie Mae and Freddie Mac by the U.S. government in 2008. The initial high entropy score thus signaled an important shift in the news.

  This distributional shift can also be illustrated through the string of words "Fannie Mae and Freddie Mac growth will be" from another article in January 2007. The actual next word was "muted." The word clouds in Figure 5 compare the conditional distributions of the next word as calculated in December 2006 (left panel) and December 2008 (right panel). The onset of the global financial crisis in between these dates is clearly reflected in the word clouds, where the sizes of words reflect their relative likelihoods under the models estimated at each date.

- In late November 2010, there were high entropy articles discussing Ireland's financial strains. In one such article, the phrase "the Euro falls to a 5 week low on growing concerns Ireland will be forced to" was followed by the word "default." But "default" had low conditional probability under the October 2010 model, thus giving the sentence high entropy. The right panel of Figure 4 shows how the conditional probability of "default" changed over time, and became much higher by 2012 and again in 2014. In this example, as in the previous examples, the initial increase in entropy signaled a change in the distribution of text associated with economic developments.

[13] The conditional probability is calculated by feeding the embeddings of the preceding string of words (e.g., "lack of progress in reining in mortgage lenders Fannie Mae and Freddie Mac leaves the economy at") into the RNN model from the prior month (e.g., December 2006) and then evaluating the probability at the target word (e.g., "risk").

- Lastly, the *ENT* peak in October 2014 we observed in Figure 2 coincides with the Fed's ending of QE3. Many moderate to high entropy articles from that month mentioned QE.

4 Time Series of Returns

In this section, we present evidence that entropy forecasts future market returns, and we compare entropy with other predictors from the literature.

4.1 In-Sample Regressions

We estimate a variety of time-series forecasting regressions of the form

$$R_{t+1,t+12} = \beta_0 + \beta_{ENT} ENT_t + \gamma^\top Control_t + \epsilon_t \tag{6}$$

where $R_{t+1,t+12}$ is the cumulative market return from month $t+1$ to month $t+12$ and ENT_t is the entropy measure in month t. $Control_t$ contains sentiment (*SEN*), squared implied volatility (*VIX2*), interest rates (*DGS10* and *DGS10-2*), the dividend yield (*DY*), the difference between actual consumption and the consumption level predicted by wealth and income (*CAY*), the market return of the previous month (*Return1*), the cumulative market return of the previous 12 months excluding the most recent month (*Return12*), the cumulative returns of the Fama-French five factors (Fama and French 2015) and momentum over the previous 60 months (*Return60*, *SMB60*, *HML60*, *RMW60*, *CMA60*, and *UMD60*). As explained in Section 2, using the 60-month lagged returns converts factors to state variables, as suggested in Maio and Santa-Clara (2012). We also use *1/CAPE*, the inverse of the cyclically adjusted price-to-earnings ratio, as a forecasting variable because, together with the rate variables, these span the excess earnings yield forecasting variable from the Fed model proposed in Maio (2013).

Column (1) of Table 3 reports the coefficient estimates (and t-statistics) for Equation (6), after scaling each coefficient estimate by the standard deviation of the independent variable.[14] A one standard deviation increase in entropy predicts a 2.821% decrease in the 12-month ahead cumulative market return. Column (2) replaces *ENT* in Equation (6) with *EPU*. The results show that *EPU* positively predicts future market returns,

[14] Table A1 in the Online Appendix shows the unscaled coefficient estimates for Equation (6). Table 3 shows that *CAY* forecasts 12-month ahead market return negatively, whereas Lettau and Ludvigson 2001 find that *CAY* forecasts returns positively. We verified that, as a standalone forecaster, *CAY* forecasts return positively when restricted to the pre-2002 period but negatively in the full sample.

which is consistent with Brogaard and Detzel (2015). Column (3) includes both *ENT* and *EPU* in a single regression model. After controlling for these other measures *ENT* still forecasts future market returns, and the standardized coefficient for *ENT* (-2.607%) is nearly unchanged from column (1).

The full-sample analysis incorporates data from the COVID-19 period, which raises the concern that this time period may disproportionately impact our results. To address this concern, we replicate the analyses detailed in the first three columns using pre-COVID-19 data, which are presented in columns (4)–(6). In the pre-COVID-19 analysis, entropy is calculated up to the end of 2018 and used to predict through the end of 2019. We find that *ENT* consistently and negatively forecasts 12-month ahead cumulative returns: a one standard deviation increase in entropy is associated with a 2.7% year-ahead negative market return. Conversely, *EPU* does not significantly predict market returns once the COVID-19 period is dropped from the analysis.

4.2 Out-of-Sample Regressions

We saw in the previous section that an increase in entropy negatively forecasts future market returns in in-sample regressions. We now examine whether entropy provides out-of-sample predictability.

Let $R_{t+1,t+12}$ denote the cumulative return over months $t+1, \cdots, t+12$. For a fixed training window of length h, for each t, we run the univariate regression

$$R_{t-i-11,t-i} = \alpha_t + \beta_t ENT_{t-i-12} + \epsilon_{t-i-12}, \quad i = 1, 2, \dots, h, \tag{7}$$

estimating $\hat{\alpha}_t$ and $\hat{\beta}_t$ from the h monthly observations in the training window (we discuss the choice of h momentarily). Then we predict the future return $R_{t+1,t+12}$ as $\hat{R}_{t+1,t+12} = \hat{\alpha}_t + \hat{\beta}_t ENT_t$. The out-of-sample R-squared is calculated as

$$1 - \frac{MSE(R_{t+1,t+12} - \hat{R}_{t+1,t+12})}{MSE(R_{t+1,t+12} - \frac{1}{h}\sum_{i=1}^{h} R_{t-i-11,t-i})} \tag{8}$$

where the numerator is the mean square error of predicting with the model in (7), and the denominator is the mean square error based on using past historical means to forecast future returns. In other words, we are comparing the prediction accuracy of a model which uses ENT_t to make a conditional return forecast versus a model which just uses the rolling mean over the past h months. If predictions using entropy are better than the rolling mean predictions, we will get a positive out-of-sample R-squared. If predicting with entropy is worse than using historical means, we will get a negative out-of-sample R-squared.

Table 4 shows the out-of-sample R-squared of entropy as well as all other control variables we used in the in-sample analysis. Each variable is used separately in univariate regressions of the form in (7). Each column represents a different choice of window length $h \in \{12, 15, 18, 21, 24\}$ months, for estimation of the forecasting model in (7). In the first row, we see that entropy produces a positive out-of-sample R-squareds in four out of the five estimation windows, and the out-of-sample R-squareds range from 0.024 (with $h = 15$ months) to 0.051 (with $h = 21$ months). In the second row, we see that *ENT_NEWS*, which captures the part of entropy due to news innovation while holding the text model fixed (we formally define *ENT_NEWS* in Section 6) produces a positive out-of-sample R-squareds in all five estimation windows, with the out-of-sample R-squareds ranging from 0.060 (with $h = 24$ months) to 0.096 (with $h = 18$ months). In fact, this news innovations measure outperforms *all* other predictors, most of which have negative R-squareds in all estimation windows. Only two other measures have non-negative out-of-sample R-squareds: *Return60* produces positive R-squareds in all estimation windows and *RMW60* produces positive R-squareds for estimation window lengths of 18, 21, and 24 months.

Table A4 in the Online Appendix shows the average of β_t from (7) for each of the training horizons used in the out-of-sample forecasting. The average forecasting coefficients associated with entropy are all negative and are highly statistically significant for horizons of 15 months and above. Though the β_t's vary over time, the average rolling entropy forecasting coefficient for future returns is consistent with the full sample results in Table 3.

Given the difficulty of out-of-sample market return forecasting (Welch and Goyal 2008, Campbell and Thompson 2008), the performance of *ENT* and *ENT_NEWS* in our out-of-sample tests is remarkable.

5 Cross-Section of Returns

According to the argument in Maio and Santa-Clara (2012), the sign with which an ICAPM state variable forecasts aggregate market returns should match the cross-sectional risk price associated with that variable. If entropy negatively forecasts the return of the aggregate stock market, then high entropy is associated with an unfavorable investment opportunity set for investors. In this case, securities that do well during high entropy times provide a useful hedge, and should therefore earn a negative risk premium. In this section, we investigate how exposure to entropy is priced in the cross-section of stock returns.

In examining how entropy is priced in the cross-section of expected stock returns, we want to test whether entropy is a priced risk factor and estimate the price of aggregate entropy risk. According to the Merton (1973) ICAPM framework, in equilibrium, an asset's risk premium is determined by its conditional variances with ex-post market returns and innovations of state variables of the form

$$\mathbb{E}_t[r_{i,t+1} - r_{f,t+1}] = \beta Cov_t(r_{i,t+1}, r_{m,t+1}) + \gamma Cov_t(r_{i,t+1}, \Delta x_{t+1}) \tag{9}$$

where $r_{i,t+1}$ is the ex-post return on asset i, $r_{f,t+1}$ is the risk-free rate, $r_{m,t+1}$ is the ex-post market return, and x_t is state variable that affect the investment opportunity set. The coefficients β and γ are risk premia. Equation (9) – which reproduces equation (8) of Maio and Santa-Clara (2012) – states that investors receive compensation for the covariance of stock returns with variables that impact future market returns. Based on the argument of Maio and Santa-Clara (2012), if Δx_t forecasts stock returns positively, then an ICAPM-type argument suggests that stocks that positively covary with Δx_t should earn positive risk premia because they add to investor risk. On the other hand, if, like a change in entropy, Δx_t forecasts future negative stock returns, then stocks that covary positively with Δx_t hedge against unfavorable changes in the investment opportunity set and should earn negative risk premia. The ICAPM sign property of Maio and Santa-Clara (2012) can be summarized as:

$$\text{sgn}(\gamma) = \begin{cases} + & \text{if } \Delta x_t \text{ predicts market returns positively,} \\ - & \text{if } \Delta x_t \text{ predicts market returns negatively.} \end{cases} \tag{10}$$

We have seen that entropy forecasts a decline in market returns, so we expect it to carry a negative risk premium.

5.1 Factor-Mimicking Portfolios

Because *ENT* and *EPU* are not directly tradeable, we approximate them with factor-mimicking portfolios, as follows. In the first step, we extract innovations from the *ENT* and *EPU* time series by fitting an $AR(p)$ model. The number of lags in the AR model is determined by minimizing the Bayesian information criterion (BIC). The final models are $AR(12)$ for ENT and $AR(1)$ for EPU:

$$\begin{aligned} ENT_t &= \beta_{ENT,0} + \sum_{i=1}^{12} \beta_{ENT,i} ENT_{t-i} + \gamma_{ENT} MKT_{t-1} + \epsilon_{ENT,t} \\ EPU_t &= \beta_{EPU,0} + \beta_{EPU,1} EPU_{t-1} + \gamma_{EPU} MKT_{t-1} + \epsilon_{EPU,t} \end{aligned} \tag{11}$$

Following Breeden et al. (1989), Lamont et al. (2001), Ang et al. (2006), in the second step, we create the mimicking factor F_{ENT} (F_{EPU}) to track innovations in ENT (EPU) by estimating the coefficient b_{ENT} (b_{EPU}) in the following regressions

$$\begin{aligned} \hat{\epsilon}_{ENT,t} &= a_{ENT} + b_{ENT}^{\top} X_t + u_{ENT,t} \\ \hat{\epsilon}_{EPU,t} &= a_{EPU} + b_{EPU}^{\top} X_t + u_{EPU,t} \end{aligned} \tag{12}$$

where $\hat{\epsilon}_{ENT,t}$ and $\hat{\epsilon}_{EPU,t}$ denote the innovations of ENT and EPU from (11), and X_t denotes the excess returns on a set of base assets. We estimate (12) over the full sample and construct factor mimicking portfolios as follows

$$\begin{aligned} F_{ENT,t} &= \hat{b}_{ENT}^{\top} X_t \\ F_{EPU,t} &= \hat{b}_{EPU}^{\top} X_t \end{aligned} \tag{13}$$

where $\hat{b}_{ENT}$ and $\hat{b}_{EPU}$ are the coefficient estimates from (12). Since X_t represents the excess returns of base assets portfolios, the coefficient vectors $\{b_{ENT}, b_{EPU}\}$ can be interpreted as weights of a zero-cost portfolio with returns given by (13).

We consider four sets of base assets: the 25 Fama-French portfolios sorted by size and momentum, size and book to market, size and investment, and size and profitability, respectively. In the risk premia tests of the next section, the base assets are selected to correspond to the test assets in the Fama and MacBeth (1973) analysis. Table 5 shows summaries of the Fama and French (2015) factor and momentum returns, as well as those of the ENT and EPU factor replicating portfolios using different base assets.

5.2 Factor Risk Premia

To estimate the risk premium associated with the entropy factor mimicking portfolio, we run a Fama and MacBeth (1973) analysis using MKT (market minus risk-free), SMB, HML, RMW, CMA, UMD, as well as F_{EPU} and F_{ENT} as the factors. To properly reflect estimation risk from the first stage regressions, we use the following general method of moments (GMM) system in our estimation

$$\begin{aligned} \mathbb{E}[r_{i,t} - \alpha_i - \beta_i^\top f_t] &= 0 \quad i = 1, 2, \cdots, N \\ \mathbb{E}[(r_{i,t} - \alpha_i - \beta_i^\top f_t) f_t] &= 0_K \quad i = 1, 2, \cdots, N \\ \mathbb{E}[\beta(r - \beta^\top \lambda)] &= 0_K \end{aligned} \tag{14}$$

where $r_{i,t}$ is the excess return of test asset i on day t, β_i is a K by 1 vector of factor loadings for test asset i, f_t is a K by 1 vector of factors, and 0_K is a K by 1 vector of zeros. In addition, $\beta = [\beta_1 \ \beta_2 \ \cdots \ \beta_N]$ is the $K \times N$ matrix of the N test asset betas, r is a vector of the N test asset average excess returns, and λ is a K by 1 vector of factor risk premia.[15] The parameters of the system are $\{\alpha_1, \cdots, \alpha_N, \beta_1, \cdots, \beta_N, \lambda\}$. The first two moment conditions, which correspond to the first-stage Fama-MacBeth regressions, exactly identify α and β. The last moment condition represents the second stage regressions, with no constant, which pin down the prices of risk. See Cochrane 2009 page 241 for details.

Table 6 shows the risk premia from (14), but scaled by the standard deviations of the first-stage betas to highlight the magnitudes of the effects.[16] Columns (1) – (4) show monthly risk premia using different sets of 25 Fama-French portfolios as the base and test assets: size-momentum, size-book/market, size-investment, and size-profitability. We see that of the four sets of test assets, F_{ENT} carries a negative and significant risk premium in all cases, with the monthly risk premium for a standard deviation increases in entropy beta ranging from -0.060% to -0.091%. The magnitude of the entropy risk premium is on par with that of the other factors, suggesting the risk compensation associated with hedging entropy risk is economically large.

In comparison, F_{EPU} carries a negative and significant risk premium in two out of the four tests as well, with a risk premium ranging from -0.063% to -0.083% per unit of standard deviation of EPU betas. The sign and the magnitude of the F_{EPU} risk premium are consistent with the results in Brogaard and Detzel (2015) (compare Panel B of their Table 5 against the unscaled risk premia for F_{EPU} in Table A5 of the Online Appendix). Among the five factors of Fama and French (2015), only MKT has a significant risk premium for more than two out of four test assets, with a risk premium ranging from 0.048% to 0.083%. The momentum factor carries a positive and significant risk premium in two tests: 0.083% for size and book-to-market portfolios and 0.064% for size and

[15] The risk premia λ in (14) differ from γ in (9) in that they are premia on regression coefficients rather than covariances.

[16] Table A5 in the Online Appendix shows the raw risk premia.

investment portfolios.

Column (5) of Table 6 shows the sign of the ICAPM property from (10). The determination of whether a factor forecasts year-ahead market returns positively or negatively is based on the forecasting regression in (6), shown in column (3) of Table 3. A "+" or "–" in column (5) of Table 6 corresponds to significant positive and negative coefficients at 5% level, respectively; a blank indicates a lack of significance in the forecasting regression. The ICAPM sign property posits that any significant coefficients in columns (1) – (4) of Table 6 should have the same signs as the corresponding entry in column (5).[17] *ENT* is the only factor for which this property holds. *CMA* has positive signs in column (5) but only one out of the four tests shows a significant risk premium.

5.3 Which Securities Hedge Entropy Risk

In light of the price of entropy risk results from the prior section, we now investigate which types of securities are useful for entropy risk hedging. We use decile portfolios from univariate sorts on size, book/market, operating profitability, and investment obtained from Ken French's website as test assets, and estimate their loadings on F_{ENT}. In particular, we estimate each asset i's factor exposures using a full-sample time-series regression:

$$r_{i,t} = \beta_{i,0} + \beta_{i,F_1} F_{1,t} + \cdots + \beta_{i,F_8} F_{8,t} + \epsilon_{i,t}$$

where $r_{i,t}$ is the excess return of asset i on day t, $F_{i,t}$ is the day-t return on factor i, and the indices $j = 1, \ldots, 8$ refer to *MKT*, *SMB*, *HML*, *RMW*, *CMA*, *UMD*, F_{EPU}, and F_{ENT}, respectively. For this analysis, F_{EPU} and F_{ENT} were constructed using 100 base assets, corresponding to the 25 size-X sorted portfolios for $X \in$ {momentum, book/market, investment, profitability}. Each β_{i,F_j} is asset i's loading on factor j. From this set of equations, we compute estimates $\hat{\beta}_{i,F_1}, \ldots, \hat{\beta}_{i,F_8}$ of the factor loadings for each asset i.

In Figure 6, we show the F_{ENT} loadings ($\hat{\beta}_{i,F_8}$) corresponding to each test asset. The first panel shows these loadings for size-sorted portfolios. The U-shaped pattern indicates that small and large companies hedge entropy risk, while medium-sized firms are negatively correlated with entropy innovations. Companies sorted on momentum (left panel, middle row) show that past losers hedge entropy risk while past winners do poorly during high entropy times, though the effect is non-monotonic. Companies sorted on book/market (right panel, middle row) show a similar pattern of exposures to the size-

[17]Recall that *UMD60* in Table 3 corresponds to *UMD* in Table 6, *Return60* corresponds to *MKT*, and so on. *ENT* and *EPU* are the state variable equivalents of F_{ENT} and F_{EPU}.

sorted portfolios, with growth firms (low book/market) and value firms (second highest book/market) having positive entropy betas, and all other portfolios having negative entropy betas.[18] In contrast, sorting on investment (left panel, bottom row) yields an inverted U-shape, indicating that very low and very high investment companies do poorly during times of high entropy, while medium-investment firms do well. The bottom right panel shows that companies with high operating profitability offer a hedge to entropy risk, while less profitable companies are entropy-risky. As a robustness check, Figure A4 in the Online Appendix repeats the analysis in each panel of Figure 6 but using F_{ENT} constructed from different base assets. The results are similar.

6 Channels

We've shown that entropy negatively and robustly forecasts aggregate market returns both in- and out-of-sample. Furthermore, entropy is a priced cross-sectional risk and is the only factor in our analysis that satisfies the Maio and Santa-Clara (2012) ICAPM sign property between its forecasting direction for the market (negative) and its cross-sectional price of risk (also negative). In this section, we investigate some potential mechanisms underlying these findings. First, we show that entropy can be decomposed into a news innovation component and a model innovation component, and that the news innovation part plays a large role in our findings. Second, we study how entropy and its different components forecast macroeconomic outcomes, and speculate on how this translates to market return predictability.

6.1 Entropy Decomposition

We decompose the entropy measure from Equation (5) into two parts

$$ENT_t = ENT\_NEWS_t + ENT\_MODEL_t. \tag{15}$$

The first part is

$$ENT\_NEWS_t = m_{[t-18,t-13]}(t) – m_{[t-18,t-13]}(t-12)$$

[18]We interpret the highest book/market decile as distressed, rather than value, stocks with extreme high book values relative to market values.

which uses the text model from one year prior to month t, but applies it to month t news flow. This captures the change in the information content of news. The second part is

$$ENT\_MODEL_t = m_{[t-6,t-1]}(t) - m_{[t-18,t-13]}(t)$$

which looks at the difference between month t entropy as seen through the lens of today's and the year prior's text models. In order to understand what drives the predictive power of *ENT*, we replace ENT_t by ENT_NEWS_t and ENT_MODEL_t, and repeat the time series prediction of 12-month ahead cumulative returns in-sample as in Section 4.1 and out-of-sample as in Section 4.2.

Online Appendix Table A2 columns (1) and (2) show the in-sample results, where the coefficient estimates are normalized by the standard deviation of the independent variables.[19] Column (1) indicates that a standard deviation increase in *ENT_NEWS* is associated with a -1.433% market return over the next 12 months. A standard-deviation increase in *ENT_MODEL*, column (2), is associated with a -1.393% next 12-month return. However, neither effect is significant.

Table 4 presents the out-of-sample R-squareds for *ENT_NEWS* and *ENT_MODEL*. *ENT_NEWS* produces a positive out-of-sample R-squared irrespective of the model estimation window length, with the largest R-squared of 0.096 occurring at an estimation window length of 18 months. On the other hand, *ENT_MODEL* never produces a positive out-of-sample R-squared. The predictive power of *ENT* for future market returns comes from the change in the information content of news rather than from the year-over-year change in the model. We return to this finding in Section 6.2.

While *ENT_NEWS* is a very robust forecaster of year-ahead market returns, the direction of forecastability fluctuates over time. This can be seen by the negative, but insignificant average β_t coefficient from (7) for *ENT_NEWS* shown in Table A4 of the Online Appendix. While the average direction of forecastability from *ENT_NEWS* to year-ahead market returns is negative, there are also times when this sign is positive, which renders the average impact of the effect only marginally significant.

Next, we explore whether *ENT_NEWS* and *ENT_MODEL* are priced cross-sectionally. We construct factor-mimicking portfolios for these two components of *ENT* using the procedure from Section 5.1. The base assets for the factor mimicking portfolios and the test assets are the 25 Fama-French portfolios: size-momentum, size-book-to-market, size-investment, and size-profitability. For all tests, the base assets are the same as the test

[19] Online Appendix Table A3 shows the unscaled results.

assets.

Table 7 shows that F_{ENT_NEWS} has a negative risk premium in three out of the four cases, ranging from -0.055% to -0.060% per month.[20] Though it is negative in all cases, the risk premium is significant for the book-to-market, investment, and profitability sorts, but not for momentum. The magnitudes of the risk premia are again large in comparison to those associated with the other factors, and are consistent with the results in Table 6, where we use F_{ENT} instead of F_{ENT_NEWS}. Column (5) shows that none of the factors satisfy the ICAPM property in (10) when the signs of the coefficients in the forecasting regression in (6) are obtained from column (1) of Table A3. However, as Tables A3 and A4 show, the general predictive sign of *ENT_NEWS* for future market returns is negative, though marginally significant.

Tables A7 (scaled coefficients) and A8 (unscaled) of the Online Appendix show that *ENT_MODEL* is not associated with a significant risk premium. And this part of entropy does not forecast future market returns in- or out-of-sample. The evidence points to the market demanding a negative risk premium for securities that help hedge the part of entropy – *ENT_NEWS*, or news innovation – which is useful for forecasting market returns in our out-of-sample tests.

6.2 Entropy and Macro Outcomes

Having established that the predictive power of *ENT* stems largely from year-over-year news innovation, we now explore whether such changes in news flow predict economic fundamentals. We estimate time-series forecasting regressions of the form

$$Y_{t+12} = \beta_0 + \beta_{ENT} ENT_t + \gamma^\top Control_t + \beta_Y Y_t + \epsilon_t. \tag{16}$$

The dependent variable[21] in our analysis is one of:

[20] Tables A6 of the Online Appendix shows the unscaled results.

[21] 12-month real earnings per share inflation-adjusted, constant May 2023 dollars.

Variable	Description	Units
EPU	economic policy uncertainty	standard units
UNRATE	unemployment rate	percent
INDPRO_YOY	year-over-year change in industrial production	percent
CPI_YOY	year-over-year change in the Consumer Price Index	percent
DGS10, *DGS2*	10-year and 2-year Treasury yield	percent
DGS10-2	10-year minus 2-year slope	percent
VIX	CBOE Volatility Index	percent
EPS	S&P500 last-twelve-month earnings per share	dollars

For each dependent variable, we use the lagged value of entropy (ENT_t), the lagged value of the variable in question, and the lagged values of all other dependent variables as controls.

Table 8 summarizes the results. Each column of Table 8 corresponds to one of the above dependent variables. The first set of rows of the table shows the β_{ENT} coefficient from (16), scaled by the full-sample standard deviation of entropy, and the associated t-statistic. The second set of rows shows the results of (16) where ENT_t is replaced with the news innovation part of entropy, ENT_NEWS_t. The third set of rows shows the results for ENT_MODEL_t.[22]

The table shows that the forecasting results for *ENT* and *ENT_NEWS* are very similar. An increase in either measure positively forecasts the unemployment rate and the *VIX* volatility index, and negatively forecasts industrial production, CPI, the level of interest rates, and the S&P 500 earnings per share. There is a weak positive forecasting relationship from these two entropy measures for *EPU*, and no relationship for the 2s-10s curve. *ENT_MODEL* does not significantly forecast any of the macro outcomes variables, with the exception of the *VIX*, where the relationship is negative, not positive as with the other entropy measures. Overall, entropy and its news innovation component robustly forecast negative economic outcomes one year ahead.

6.3 Discussion

As in the return forecasting regressions, high *ENT* and high *ENT_NEWS* significantly forecast negative macroeconomic outcomes. On the other hand, the model innovation part

[22] The full results reporting the scaled coefficients are in Online Appendix Tables A9–A11. The full results reporting the raw coefficients are in Online Appendix Tables A12–A14.

of entropy (*ENT_MODEL*) does not forecast macroeconomic outcomes. Furthermore, exposure to both *ENT* and *ENT_NEWS* receives a significant and negative cross-sectional risk premium, while *ENT_MODEL* is not a priced risk. The evidence points to a somewhat puzzling finding. Market participants understand that there is predictability from entropy (newness of news) to future market and macroeconomic outcomes because they are willing to give up expected returns to hold securities that allow them to hedge this risk. However, the aggregate market does not react to this information instantaneously, which is why entropy and its news innovation component forecast year-ahead market returns. One may have expected entropy to forecast aggregate market returns positively, as market participants demand compensation to hold risky assets heading into difficult economic periods. However, we do not find this result. Instead, our finding of aggregate underreaction is consistent with recent evidence of asset-class-level macro momentum (Brooks et al. 2023).

The non-instantaneous reaction of markets to entropy may have two causes. First, consistent with the rational inattention theory of Sims (2003), market participants may not have been aware of times of high news entropy. Since our entropy measure is difficult to observe – depending on tens of thousands of articles today and one year ago, as well as sophisticated NLP techniques – it is possible that, historically, market participants were not immediately aware of high entropy times. Even if such times were only identifiable after the fact, securities that did well in these ex-post high entropy periods (see discussion in Section 5.3), may have been bid up by investors in anticipation of similar hedging properties in the future. Thus it is possible to observe cross-sectional entropy risk premia even if market participants could not observe entropy in real-time. With advances in computational linguistics and data availability, this channel may or may not be as relevant in the future.

The second, arguably more enduring, reason for such underreaction, is a combination of slow-moving institutional capital (Gabaix and Koijen 2021, Glasserman et al. 2023) with limits to arbitrage (Gromb and Vayanos 2010). Institutions may be constrained by mandate to not deviate excessively from a particular set of portfolio targets, for example, a risk level consistent with a 60/40 stock/bond portfolio. Even if such institutions begin to believe that markets will not do well over the next year, the amount of uncertainty around that belief is still large (recall that even the highest out-of-sample R-squareds for year-ahead market returns in Table 4 are less than 10%). Such institutions may need board or investment consultant approval before they can change their portfolio allocations, and these approvals may be slow in coming. Arbitrage capital, like actively managed mutual funds or hedge funds, may not be big enough relative to the amount of capital controlled by constrained institutions to provide a sufficient offset. As such, macro-level information may take time to work its way into stock prices. This channel is unlikely to be impacted by the availability of better or faster information, suggesting that market return forecastability by entropy may be a persistent market phenomenon.

7 Robustness

In this section, we check whether entropy is spanned by existing risk factors, whether the entropy risk premium can be explained by a news sentiment risk premium, and whether other measures of uncertainty span entropy or can explain the forecasting power of entropy for market returns.

7.1 Existing Pricing Factors

Entropy is distinctive because of its impressive in- and out-of-sample forecasting power for market returns, the fact that it is the only factor that satisfies the ICAPM sign property of (10), and its macroeconomic forecasting ability. However, it is possible that the information content of entropy is spanned by one of the multitude of existing factors in the literature. To check for this, we use the 153 value-weighted factors discussed in Jensen et al. (2022), whose daily returns are available at `https://jkpfactors.com/`. We augment this set of factors with the ENT and EPU factor replicating portfolios, F_{ENT} and F_{EPU}, using size-momentum as the base assets.[23] We regress each of the 155 factor return series on the remaining 154 factors.

In Figure 7 we show these R-squareds for each factor, ranked from low to high. Factors close to the lower left corner are not well explained by the existing factor zoo, while factors close to the upper right corner are largely spanned by existing, alternative factors. We label the top and bottom five factors based on this spanning R-squared measure and also indicate ENT by red text. F_{ENT}, the entropy mimicking portfolio with size-momentum as the base assets, has the seventh lowest R-squared among the 155 factors. This shows that ENT conveys novel information relative to existing factors.

[23]The results for the other base assets are similar.

7.2 Economic Sentiment

Another well-known news-based variable is *SEN*, the San Francisco Fed's news sentiment index. The main return forecasting results in Table 3 show that entropy is a strong forecaster of year-ahead market returns even after controlling for the information content of *SEN*, while *SEN* is not a significant return forecaster when *ENT* is included in the set of forecasting variables. Similarly, Table 4 shows that *SEN* is a poor out-of-sample forecaster of market returns. Online Appendix Tables A15 (scaled coefficients) and A16 (unscaled) show the Fama-MacBeth risk premium analysis which includes F_{SEN}, the *SEN* replicating portfolio constructed using the method described in Section 5.1, in addition to the other factor exposures. Exposure to *SEN* carries a positive and significant risk premium in two out of the four tests but does not satisfy the ICAPM property because it is not a significant in-sample market return forecaster. Even in the presence of F_{SEN}, exposure to F_{ENT} continues to carry a negative and significant risk premium, consistent with the results of Section 5.2.

7.3 Other Uncertainty Measures

Another concern is that entropy may be spanned by other uncertainty measures studied in the literature. To check for this, we utilize measures of macroeconomic uncertainty discussed in the survey article David and Veronesi (2022).[24] In particular, we consider the Bekaert et al. (2022) (*BEX*) uncertainty measure of time-varying risk aversion, the Jurado et al. (2015) (*JLN*) variable which provides econometric estimates of the conditional volatility of the purely unforecastable component of the future values of "hundreds of macroeconomic and financial indicators," the Azzimonti (2018) Partisan Conflict Index (*PCI*) which tracks the degree of disagreement among U.S. politicians at the federal level, as well as *ENT*, *EPU*, *SEN*, and *VIX*.

We regress each of the seven uncertainty measures on the other six uncertainty measures and calculate the resultant R-squared. In Figure 8 we show these R-squareds for each uncertainty measure, ranked from lowest to highest. Variables close to the lower left corner are not well explained by the remaining ones, while variables close to the upper right corner are largely spanned by the remaining measures. *ENT*, indicated by a red dot, has the lowest R-squared (0.13) among all considered uncertainty measures, while *EPU* has the highest R-squared (0.70). The second smallest is *PCI*, whose R-squared of 0.47 is still substantially greater than that of *ENT*. Our entropy measure is the one that is

[24] All the data used in this part of the analysis are available from the survey article.

furthest removed from the information content of standard measures of uncertainty.

We expand the in-sample analysis of Section 4.1 to include multiple uncertainty measures along with the more standard controls used previously in Table 3. We estimate a variety of time-series forecasting regressions of the form

$$R_{t+1,t+12} = \beta_0 + \beta_{ENT} ENT_t + \eta^\top Uncertain_t + \gamma^\top Control_t + \epsilon_t \tag{17}$$

where $R_{t+1,t+12}$ is the cumulative market return from month $t+1$ to month $t+12$ and ENT_t is the entropy measure in month t. $Uncertain_t$ contains non-text-based uncertainty measures: *BEX*, *JLN*, *PCI*, and squared implied volatility (*VIX2*). $Control_t$ contains interest rates (*DGS10* and *DGS10-2*), the dividend yield (*DY*), the difference between actual consumption and the consumption level predicted by wealth and income (*CAY*), the inverse of the cyclically adjusted price-to-earnings ratio (*1/CAPE*), the market return of the previous month (*Return1*), the cumulative market return of the previous 12 months excluding the most recent month (*Return12*), the cumulative returns of the Fama-French five factors (Fama and French 2015) and momentum over the previous 60 months (*Return60*, *SMB60*, *HML60*, *RMW60*, *CMA60*, and *UMD60*). These are the forecasting variables used in the analysis in Table 3.

Columns (1) and (2) of Table 9 report the scaled coefficient estimates (and t-statistics) for (17), without and with the uncertainty controls.[25] A one standard deviation increase in entropy predicts a 2.266%-2.530% decrease in the 12-month ahead market return. Columns (3) and (4) replace *ENT* in (17) with two text-based measures: economic policy uncertainty (*EPU*) and sentiment (*SEN*). The results show that *EPU* positively predicts future market returns but *SEN* does not have a significant impact. Columns (5) and (6) include all text-based uncertainty measures *ENT*, *EPU*, and *SEN* in a single regression model, again without and with uncertainty controls. After controlling for these other measures *ENT* still forecasts future market returns, and the standardized coefficients for *ENT* (from -2.306% to -2.413%) are nearly unchanged from columns (1) and (2). In all cases, introducing uncertainty controls (Columns 2, 4, 6) does not meaningfully change the results without the uncertainty controls (Columns 1, 3, 5).

We also replicate the analysis in Section 4.2 to test if any of the uncertainty measures are robust out-of-sample forecasters of aggregate market returns. The results for *BEX*, *JLN*, *PCI* and *VIX2* in Table 4 show that none of the uncertainty measures forecast market returns out-of-sample.

[25] The full regression results are shown in Table A17 (scaled coefficients) and Table A18 (raw coefficients) of the Online Appendix.

8 Conclusion

This paper combines natural language processing with the tools of empirical asset pricing to investigate a novel aspect of how news affects prices at the aggregate level. We use a recurrent neural network to derive entropy, a measure of the novelty or unusualness of aggregate news. This measure extends the prior literature (Glasserman and Mamaysky 2019) by applying modern NLP tools to solve the sparsity and context problems with prior entropy measures.

We show that entropy negatively forecasts next twelve-month market returns, even after controlling for a multitude of known in-sample return forecasters. In particular, entropy does better than either economic policy uncertainty or news sentiment in our in-sample tests. In an out-of-sample market forecasting horse race, we find that, remarkably, entropy is the best forecaster of year-ahead market returns out of a large set of candidate variables.

Using a Fama-MacBeth GMM framework, we show that entropy has a negative risk price in the cross-section of portfolio returns. Together with the finding that entropy forecasts market returns negatively, it turns out entropy is the *only* factor in our study that is consistent with the Maio and Santa-Clara (2012) ICAPM sign property. Entropy thus proxies for a factor that negatively impacts investors' opportunity set and investors are willing to give up expected return to hedge against entropy innovations.

We show that entropy can be decomposed into one part that reflects the change in news flow and the other part that reflects the change in the text model, and that it is the former, change-in-news, that accounts for entropy's time series forecasting properties and its cross-sectional risk pricing. We further show that the factor mimicking portfolio for entropy is among the least well-spanned out of 155 long-short factors we obtain from Jensen et al. (2022).

The forecasting power of entropy for future market returns likely stems from its ability to forecast future fundamental variables. Both entropy and its news innovation component positively forecast 12-month ahead unemployment and volatility, and negatively forecast industrial production, inflation, interest rates, and S&P500 corporate earnings. While entropy risk is priced, markets appear to not fully react to the information content of entropy either because of informational constraints or because of slow-moving institutional capital.

Our paper opens up several interesting directions for future work. First, the field of natural language processing is witnessing rapid advances which may produce more successful network architectures than recurrent neural networks for measuring news unusualness (e.g., Mikolov et al. 2010, Lee et al. 2017, Devlin et al. 2018, Xiong et al. 2019). Second, our analysis focuses on the impact of entropy at the aggregate level, while there may also be interesting applications at the level of individual stocks (Glasserman and Mamaysky 2019). Finally, our entropy factor replicating portfolio should be a useful factor for asset pricing models.

References

Aït-Sahalia Y, Matthys F, Osambela E, Sircar R (2021) When uncertainty and volatility are disconnected: implications for asset pricing and portfolio performance. Technical report, National Bureau of Economic Research.

Ang A, Hodrick RJ, Xing Y, Zhang X (2006) The cross-section of volatility and expected returns. *The Journal of Finance* 61(1):259–299.

Azzimonti M (2018) Partisan conflict and private investment. *Journal of Monetary Economics* 93:114–131.

Baker SR, Bloom N, Davis SJ (2016) Measuring economic policy uncertainty. *The quarterly journal of economics* 131(4):1593–1636.

Bekaert G, Engstrom EC, Xu NR (2022) The time variation in risk appetite and uncertainty. *Management Science* 68(6):3975–4004.

Breeden DT, Gibbons MR, Litzenberger RH (1989) Empirical tests of the consumption-oriented capm. *The Journal of Finance* 44(2):231–262.

Brenner M, Izhakian Y (2018) Asset pricing and ambiguity: Empirical evidence. *Journal of Financial Economics* 130(3):503–531.

Brogaard J, Detzel A (2015) The asset-pricing implications of government economic policy uncertainty. *Management Science* 61(1):3–18.

Brooks J, Feilbogen N, Ooi YH, Akant A (2023) Economic trend. Technical report, AQR.

Campbell JY, Shiller RJ (1988) Stock prices, earnings, and expected dividends. *The Journal of Finance* 43(3):661–676.

Campbell JY, Thompson SB (2008) Predicting excess stock returns out of sample: Can anything beat the historical average? *The Review of Financial Studies* 21(4):1509–1531.

Charniak E (2019) *An Introduction to Deep Learning* (MIT Press).

Chen A, Zimmerman T (2021) Open source cross-sectional asset pricing. Technical report, SSRN.

Cochrane J (2009) *Asset pricing* (Princeton University Press).

Cohen L, Malloy C, Nguyen Q (2020) Lazy prices. *The Journal of Finance* 75(3):1371–1415.

David A, Veronesi P (2022) A survey of alternative measures of macroeconomic uncertainty: Which measures forecast real variables and explain fluctuations in asset volatilities better? *Annual Review of Financial Economics* 14:439–463.

Devlin J, Chang MW, Lee K, Toutanova K (2018) Bert: Pre-training of deep bidirectional transformers for language understanding. Technical report, arXiv preprint arXiv:1810.04805.

Fama EF, French KR (2015) A five-factor asset pricing model. *Journal of Financial Economics* 116(1):1–22.

Fama EF, MacBeth JD (1973) Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy* 81(3):607–636.

Feng G, Giglio S, Xiu D (2020) Taming the factor zoo: A test of new factors. *The Journal of Finance* 75(3):1327–1370.

Gabaix X, Koijen RS (2021) In search of the origins of financial fluctuations: The inelastic markets hypothesis. Technical report, working paper.

Glasserman P, Li F, Mamaysky H (2023) Time variation in the news-returns relationship. Technical report, SSRN.

Glasserman P, Mamaysky H (2019) Does unusual news forecast market stress? *Journal of Financial and Quantitative Analysis* 54(5):1937–1974.

Goodfellow I, Bengio Y, Courville A (2016) *Deep learning* (MIT press).

Gromb D, Vayanos D (2010) Limits of arbitrage. *Annual Review of Financial Economics* 2(1):251–275.

Hansen LP, Sargent TJ (2008) Risk sensitivity, model uncertainty, and asset pricing. *Robustness*, 307–326 (Princeton University Press).

Harvey CR, Liu Y, Zhu H (2016) … and the cross-section of expected returns. *The Review of Financial Studies* 29(1):5–68.

Hoberg G, Phillips G (2016) Text-based network industries and endogenous product differentiation. *Journal of Political Economy* 124(5):1423–1465.

Hoberg G, Phillips GM (2018) Text-based industry momentum. *Journal of Financial and Quantitative Analysis* 53(6):2355–2388.

Hochreiter S, Schmidhuber J (1997) Long short-term memory. *Neural computation* 9(8):1735–1780.

Jensen TI, Kelly BT, Pedersen LH (2021) Is there a replication crisis in finance? Technical report, National Bureau of Economic Research.

Jensen TI, Kelly BT, Pedersen LH (2022) Is there a replication crisis in finance? *Journal of Finance, Forthcoming* .

Jiang F, Lee J, Martin X, Zhou G (2019) Manager sentiment and stock returns. *Journal of Financial Economics* 132(1):126–149.

Jurado K, Ludvigson SC, Ng S (2015) Measuring uncertainty. *American Economic Review* 105(3):1177–1216.

Jurafsky D, Martin J (2023) Speech and Language Processing, 3rd Edition.

Keskar NS, Mudigere D, Nocedal J, Smelyanskiy M, Tang PTP (2016) On large-batch training for deep learning: Generalization gap and sharp minima. Technical report, arXiv preprint arXiv:1609.04836.

Kingma DP, Ba J (2014) Adam: A method for stochastic optimization. Technical report, arXiv preprint arXiv:1412.6980.

Lamont O, Polk C, Saaá-Requejo J (2001) Financial constraints and stock returns. *The review of financial studies* 14(2):529–554.

Lee K, He L, Lewis M, Zettlemoyer L (2017) End-to-end neural coreference resolution. Technical report, arXiv preprint arXiv:1707.07045.

Lettau M, Ludvigson S (2001) Consumption, aggregate wealth, and expected stock returns. *the Journal of Finance* 56(3):815–849.

Maio P (2013) The "fed model" and the predictability of stock returns. *Review of Finance* 17(4):1489–1533.

Maio P, Santa-Clara P (2012) Multifactor models and their consistency with the icapm. *Journal of Financial Economics* 106(3):586–613.

Martin I (2017) What is the expected return on the market? *The Quarterly Journal of Economics* 132(1):367–433.

Merton RC (1973) An intertemporal capital asset pricing model. *Econometrica: Journal of the Econometric Society* 867–887.

Mikolov T, Karafiát M, Burget L, Cernockỳ J, Khudanpur S (2010) Recurrent neural network based language model. *Interspeech*, volume 2 (3), 1045–1048 (Makuhari).

Pennington J, Socher R, Manning CD (2014) Glove: Global vectors for word representation. *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, 1532–1543.

Radford A, Wu J, Child R, Luan D, Amodei D, Sutskever I, et al. (2019) Language models are unsupervised multitask learners. *OpenAI blog* 1(8):9.

Shapiro AH, Sudhof M, Wilson DJ (2022) Measuring news sentiment. *Journal of econometrics* 228(2):221–243.

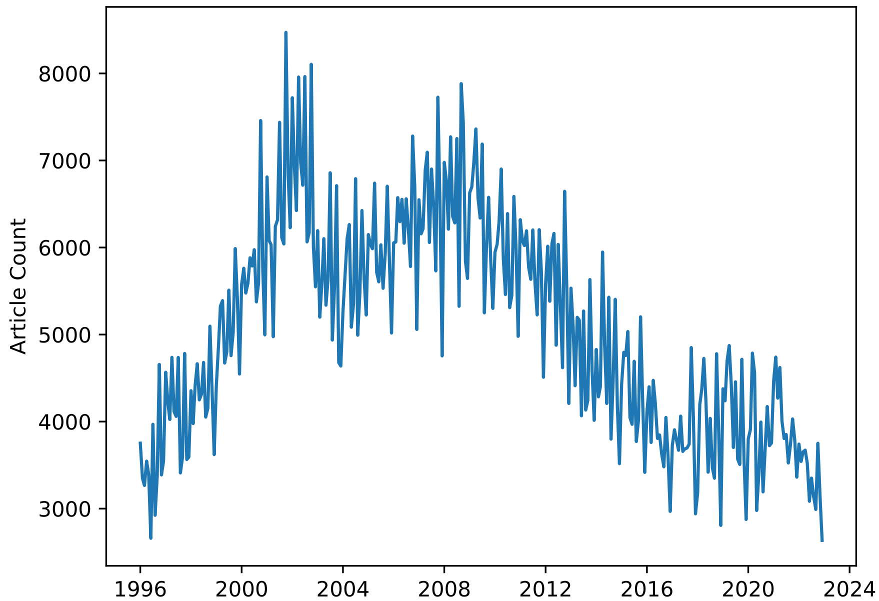
Sims CA (2003) Implications of rational inattention. *Journal of Monetary Economics* 50(3):665–690.

Tetlock PC (2011) All the news that's fit to reprint: Do investors react to stale information? *The Review of Financial Studies* 24(5):1481–1512.

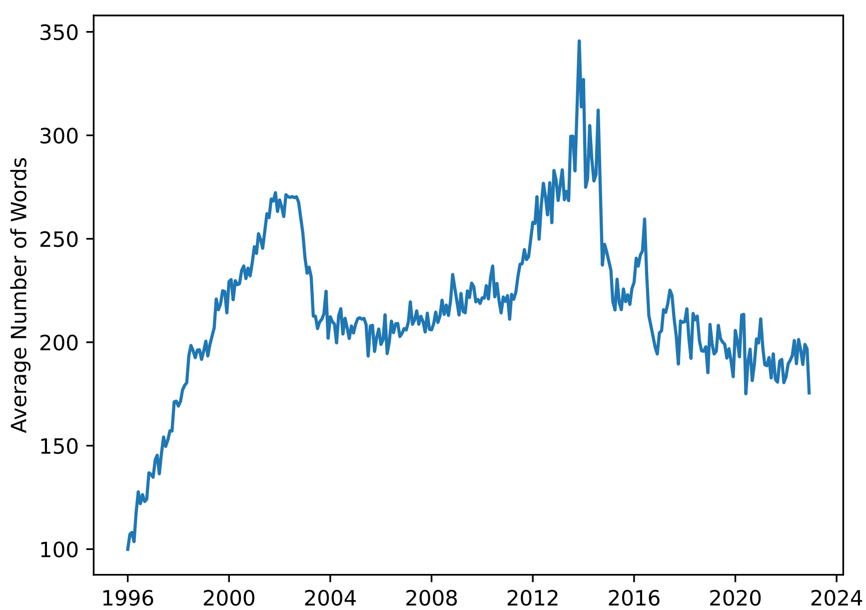
Welch I, Goyal A (2008) A comprehensive look at the empirical performance of equity premium prediction. *The Review of Financial Studies* 21(4):1455–1508.

Werbos PJ (1990) Backpropagation through time: what it does and how to do it. *Proceedings of the IEEE* 78(10):1550–1560.

Xiong W, Du J, Wang WY, Stoyanov V (2019) Pretrained encyclopedia: Weakly supervised knowledge-pretrained language model. Technical report, arXiv preprint arXiv:1912.09637.



(a) Time series of number of articles in each month.



(b) Time series of average article length in each month.

Figure 1: Thomson Reuters News Feed Direct archive articles characteristics from January 1996 to December 2022.

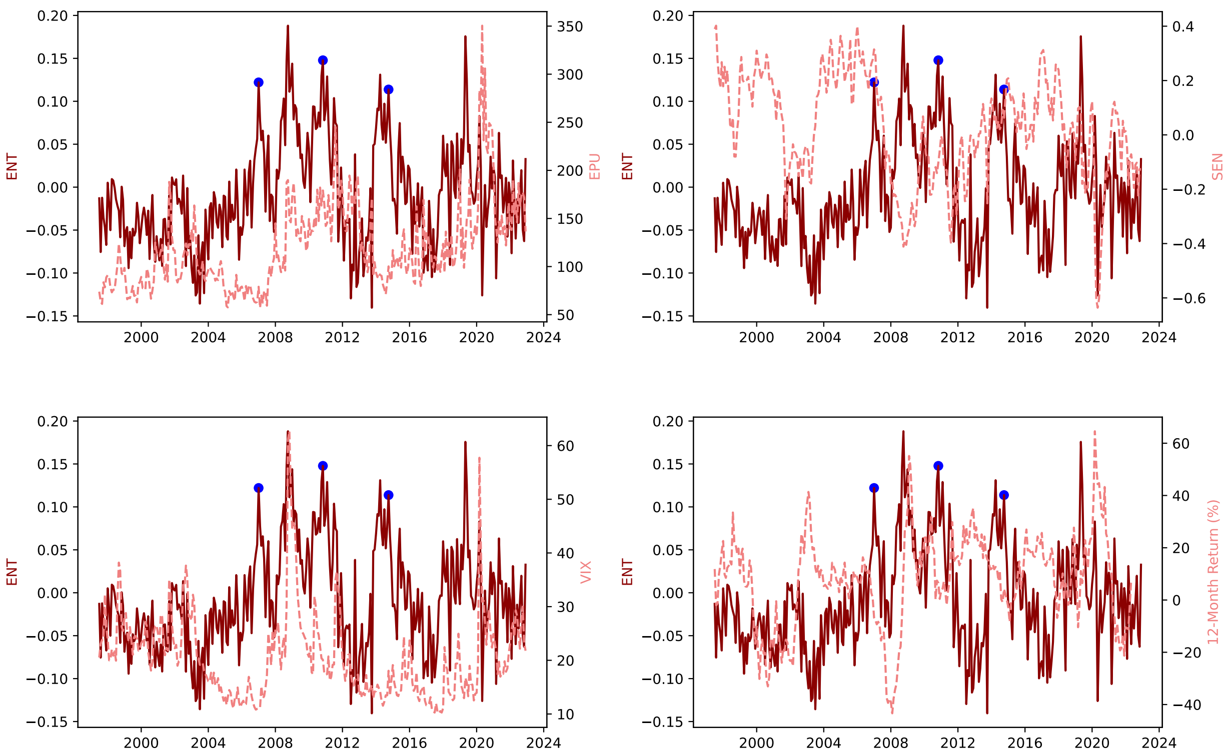


Figure 2: The top-left panel plots ENT and EPU. The top-right panel plots ENT and SEN. The bottom-left panel plots ENT and the VIX index. The bottom-right panel plots ENT and 12-month ahead cumulative returns on the CRSP value-weighted index (ends on December 2021). In all cases, ENT is plotted as the solid line, with the other series plotted as the dashed line. Data are shown on a monthly frequency. The blue dots on each series correspond to the months January 2007, November 2010, and October 2014, which are discussed in Section 3.4.

	ENT	ENT_NEWS	ENT_MODEL	EPU	SEN	VIX2	DGS10	DGS10-2	DY	CAY	1/CAPE	Return1	Return12	Return60	SMB60	HML60	RMW60	CMA60	UMD60	12-Month Ahead Return
ENT	1.000	0.859	-0.171	0.115	-0.280	0.217	-0.129	0.198	0.394	-0.217	0.186	-0.082	0.019	-0.008	0.019	-0.148	-0.061	-0.300	-0.314	-0.027
ENT_NEWS	0.859	1.000	-0.651	0.122	-0.282	0.276	-0.137	0.224	0.501	-0.164	0.186	-0.070	0.004	0.002	-0.024	-0.142	-0.158	-0.365	-0.348	0.057
ENT_MODEL	-0.171	-0.651	1.000	-0.065	0.127	-0.210	0.072	-0.138	-0.381	-0.005	-0.083	0.013	0.022	-0.016	0.075	0.053	0.213	0.259	0.204	-0.149
EPU	0.115	0.122	-0.065	1.000	-0.740	0.404	-0.598	0.171	0.319	-0.439	0.506	-0.086	-0.128	-0.165	-0.236	-0.613	0.026	-0.378	-0.298	0.415
SEN	-0.280	-0.282	0.127	-0.740	1.000	-0.525	0.451	-0.319	-0.490	0.215	-0.481	0.166	0.384	0.257	0.209	0.465	-0.110	0.341	0.227	-0.260
VIX2	0.217	0.276	-0.210	0.404	-0.525	1.000	0.032	0.089	0.359	0.191	0.034	-0.402	-0.388	-0.136	-0.238	-0.074	0.181	-0.141	0.244	0.263
DGS10	-0.129	-0.137	0.072	-0.598	0.451	0.032	1.000	-0.266	-0.442	0.741	-0.772	-0.071	-0.093	-0.011	0.063	0.599	0.150	0.386	0.576	-0.383
DGS10-2	0.198	0.224	-0.138	0.171	-0.319	0.089	-0.266	1.000	0.397	0.057	0.690	0.021	-0.012	-0.421	0.356	-0.009	0.251	0.236	-0.240	0.166
DY	0.394	0.501	-0.381	0.319	-0.490	0.359	-0.442	0.397	1.000	-0.199	0.560	-0.127	-0.338	-0.386	0.162	-0.129	0.120	-0.208	-0.396	0.488
CAY	-0.217	-0.164	-0.005	-0.439	0.215	0.191	0.741	0.057	-0.199	1.000	-0.513	-0.125	-0.211	-0.262	0.188	0.620	0.382	0.530	0.514	-0.182
1/CAPE	0.186	0.186	-0.083	0.506	-0.481	0.034	-0.772	0.690	0.560	-0.513	1.000	0.064	0.038	-0.238	0.167	-0.416	0.084	-0.115	-0.514	0.394
Return1	-0.082	-0.070	0.013	-0.086	0.166	-0.402	-0.071	0.021	-0.127	-0.125	0.064	1.000	0.035	-0.032	0.018	-0.092	-0.070	-0.044	-0.166	0.001
Return12	0.019	0.004	0.022	-0.128	0.384	-0.388	-0.093	-0.012	-0.338	-0.211	0.038	0.035	1.000	0.390	0.068	-0.184	-0.266	-0.127	-0.390	-0.025
Return60	-0.008	0.002	-0.016	-0.165	0.257	-0.136	-0.011	-0.421	-0.386	-0.262	-0.238	-0.032	0.390	1.000	-0.630	-0.378	-0.727	-0.526	-0.005	-0.287
SMB60	0.019	-0.024	0.075	-0.236	0.209	-0.238	0.063	0.356	0.162	0.188	0.167	0.018	0.068	-0.630	1.000	0.570	0.433	0.711	-0.096	-0.022
HML60	-0.148	-0.142	0.053	-0.613	0.465	-0.074	0.599	-0.009	-0.129	0.620	-0.416	-0.092	-0.184	-0.378	0.570	1.000	0.480	0.796	0.433	-0.158
RMW60	-0.061	-0.158	0.213	0.026	-0.110	0.181	0.150	0.251	0.120	0.382	0.084	-0.070	-0.266	-0.727	0.433	0.480	1.000	0.579	0.114	0.285
CMA60	-0.300	-0.365	0.259	-0.378	0.341	-0.141	0.386	0.236	-0.208	0.530	-0.115	-0.044	-0.127	-0.526	0.711	0.796	0.579	1.000	0.401	-0.032
UMD60	-0.314	-0.348	0.204	-0.298	0.227	0.244	0.576	-0.240	-0.396	0.514	-0.514	-0.166	-0.390	-0.005	-0.096	0.433	0.114	0.401	1.000	-0.204
12-Month Ahead Return	-0.027	0.057	-0.149	0.415	-0.260	0.263	-0.383	0.166	0.488	-0.182	0.394	0.001	-0.025	-0.287	-0.022	-0.158	0.285	-0.032	-0.204	1.000

Color scale: 1.00, 0.75, 0.50, 0.25, 0.00, −0.25, −0.50, −0.75, −1.00

Figure 3: Correlations at monthly observation using data from July 1997 to December 2022, inclusive.

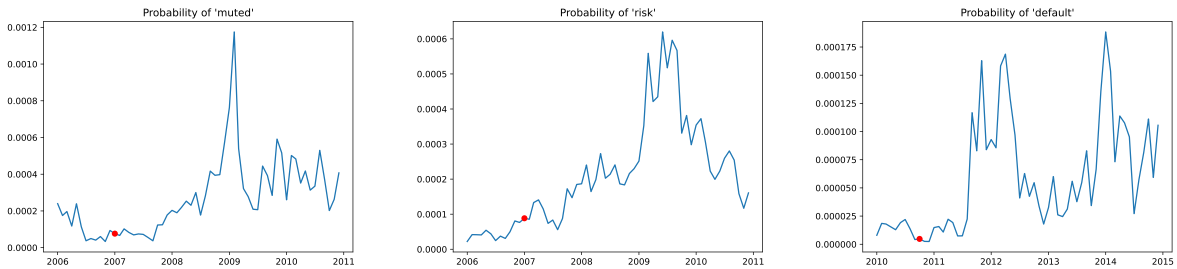


Figure 4: The left panel plots the probability of "muted" following "fannie mae and freddie mac growth will be." The middle panel plots the probability of "risk" following "lack of progress in reining in mortgage lenders fannie mae and freddie mac leaves the economy at." The right panel plots the probability of "default" following "growing concerns Ireland will be forced to." The red dot in each panel marks the date the corresponding sentence appears in the database.

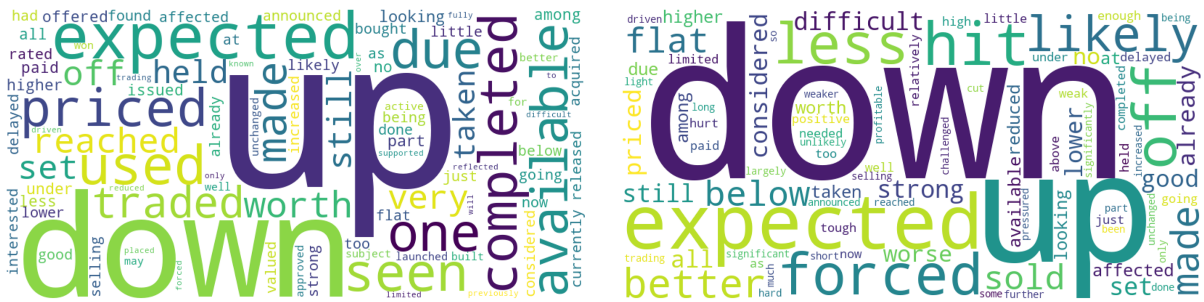


Figure 5: The most probable 100 words following the string of words "fannie mae and freddie mac growth will be" under the December 2006 model (left panel) and under the December 2008 model (right panel). The size of each word is proportional to its probability.

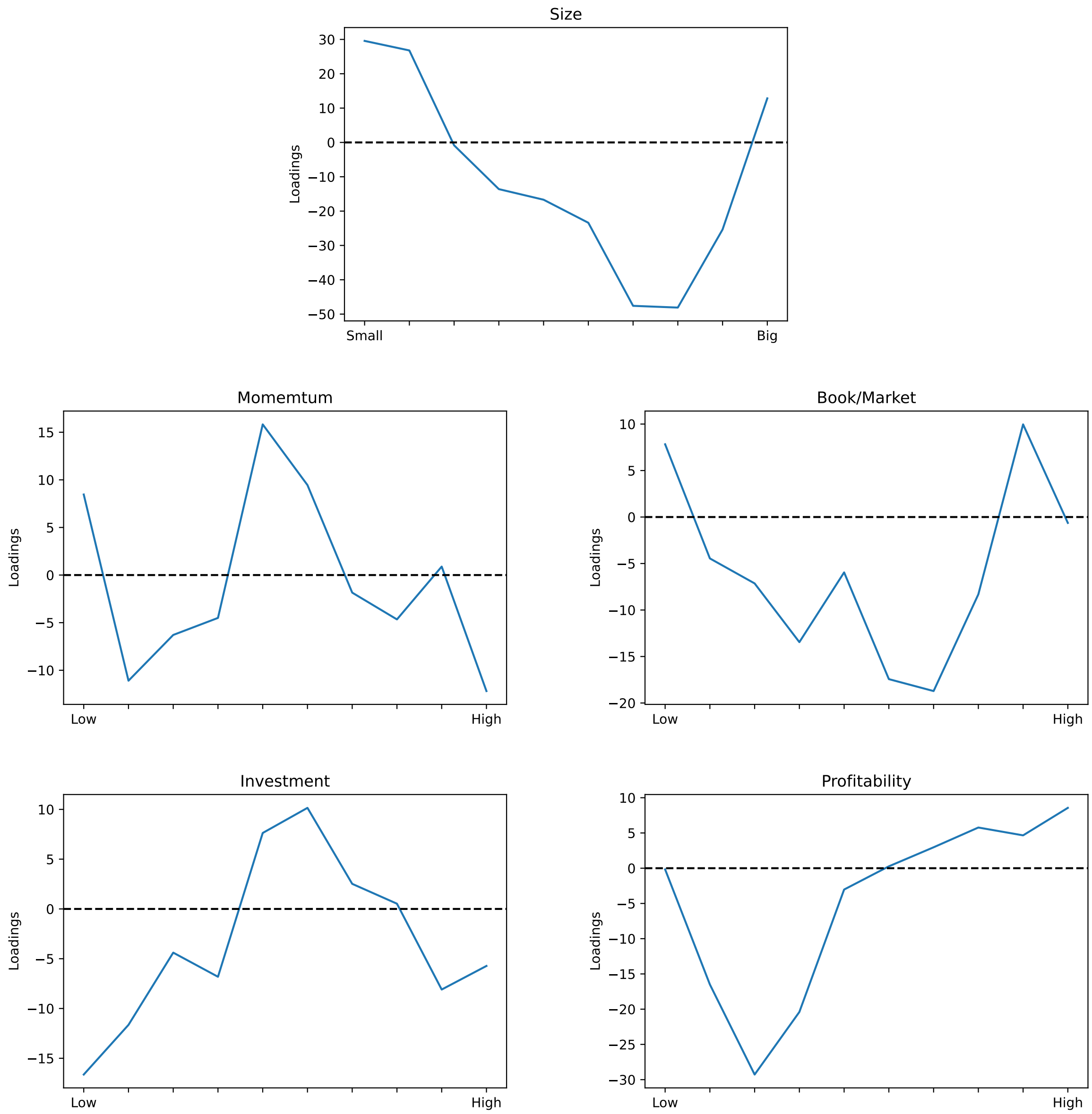


Figure 6: The first panel shows loadings on F_{ENT} of 10 portfolios sorted on size. The other panels show loadings on F_{ENT} of portfolios sorted on momentum, book/market, investment, and operating profitability, respectively. For each panel, we construct F_{ENT} using 100 base assets: 25 portfolios formed on size-momentum, size-book/market, size-investment, and size-profitability. The methodology is described in Section 5.3.

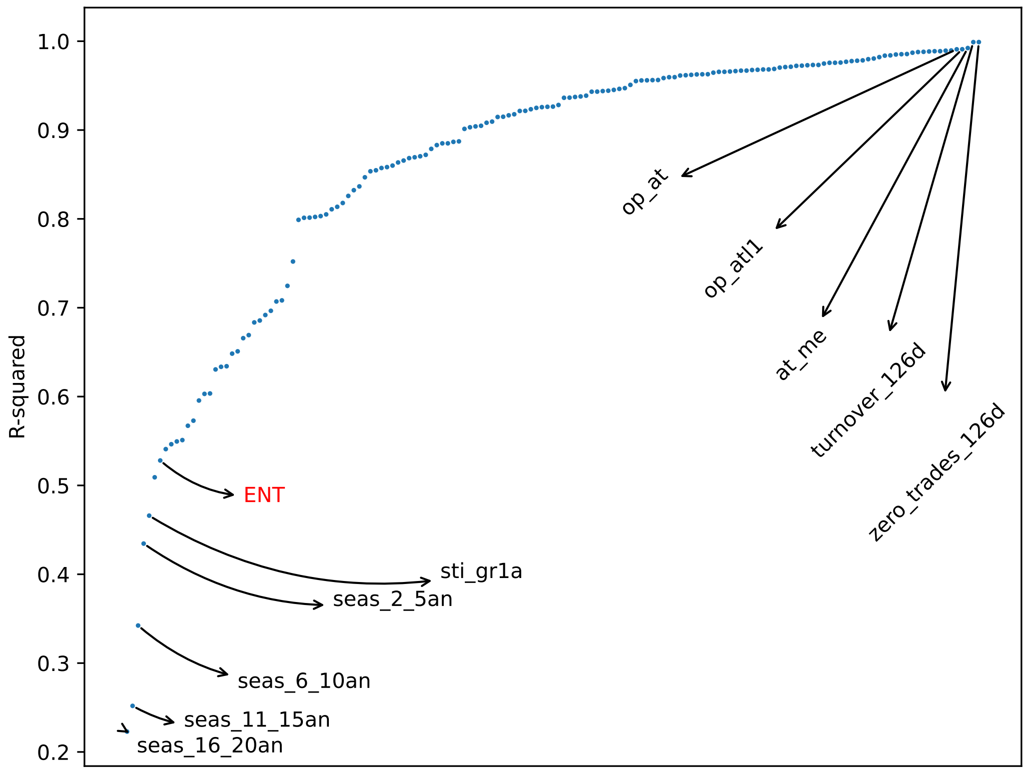


Figure 7: In-sample R^2, ordered from lowest to highest, of regressing each 153 value-weighted factors in Jensen et al. (2022) plus F_{ENT} and F_{EPU} on all other factors using the OLS model. Top and bottom five factors based on this spanning R-squared measure are annotated. Five factors with the highest R-squared are zero_trades_126d (number of zero trades with turnover as tiebreaker, 6 months), turnover_126d (share turnover), at_me (assets to market), ivol_ff3_21d (idiosyncratic volatility from the Fama-French 3-factor model), op_atl1 (ball operating profit scaled by lagged assets), and op_at (ball operating profit to assets). Five factors with the lowest R-squared are aseas_2_5an (year 2-5 lagged returns, annual), seas_6_10an (year 6-10 lagged returns, annual), sti_gr1a (change in short-term investments), seas_11_15an (year 11-15 lagged returns, annual), seas_16_20an (year 16-20 lagged returns, annual). The entropy mimicking portfolio, F_{ENT}, ranks as the seventh lowest R-squared among all 155 factors.

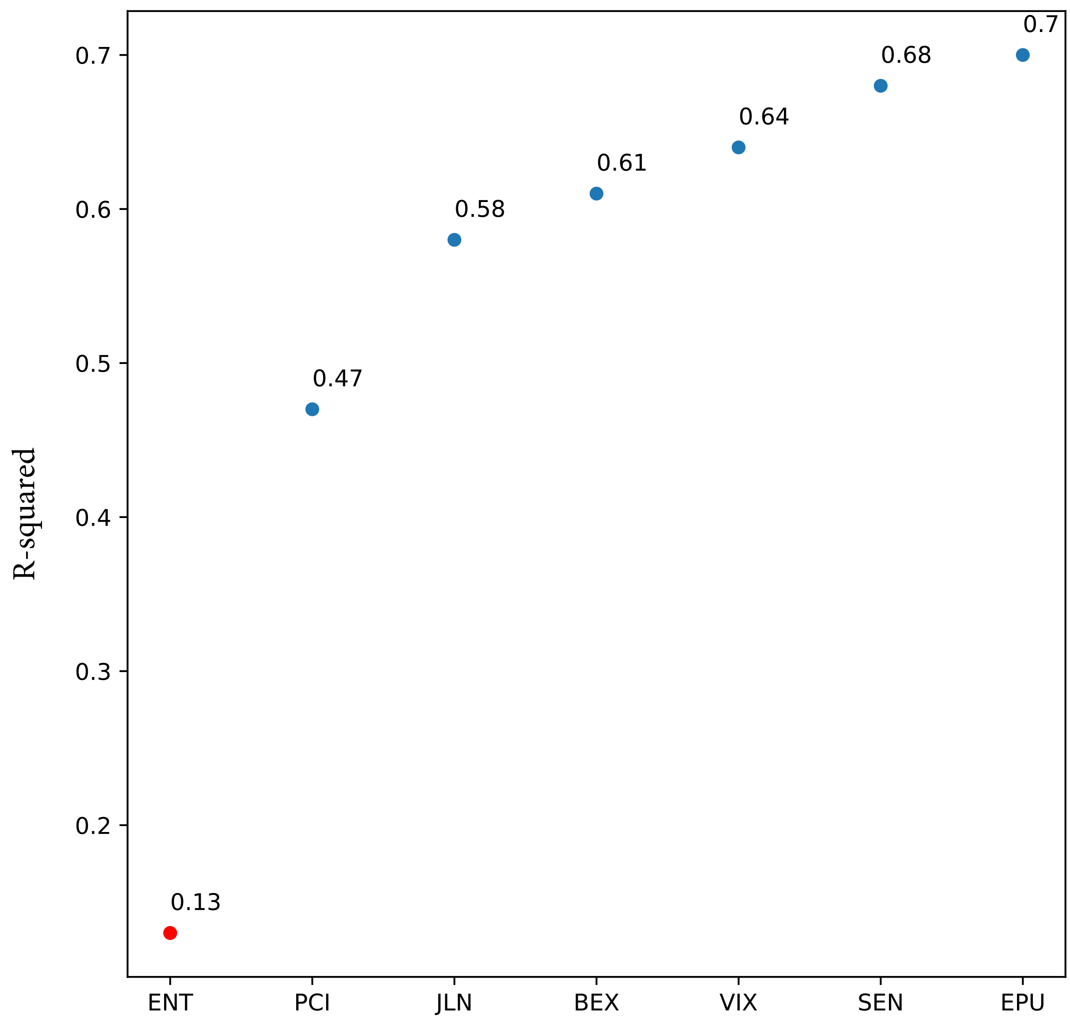


Figure 8: In-sample R^2, ordered from lowest to highest, of regressing each variable on all other variables using OLS model. Included measures of macroeconomic uncertainty, ranked from lowest to highest, are entropy (ENT, marked in red), the Azzimonti (2018) Partisan Conflict Index (PCI), the Jurado et al. (2015) (JLN) uncertainty measure, the Bekaert et al. (2022) (BEX) uncertainty measure, the Chicago Board Options Exchange's CBOE Volatility Index (VIX), the Shapiro et al. (2022) news sentiment index (SEN), and the Baker et al. (2016) Economic Policy Uncertainty (EPU).

Table 1: This table presents summary statistics of variables at the monthly frequency used in this study, including entropy measure (*ENT*), the first part of the decomposition of entropy (*ENT_NEWS*), the second part of the decomposition of entropy (*ENT_MODEL*), economic policy uncertainty (*EPU*), sentiment (*SEN*), squared implied volatility (*VIX2*), market yield on U.S. treasury securities at 10-year constant maturity (*DGS10*), market yield on U.S. treasury securities at 10-year constant maturity minus 2-year constant maturity (*DGS10-2*), the dividend yield (*DY*), the difference between actual consumption and the consumption level predicted by wealth and income (*CAY*), the inverse of the cyclically adjusted price-to-earnings ratio (*1/CAPE*), the market return of the previous month (*Return1*), the cumulative market return of the previous 12 months excluding the most recent month (*Return12*), the cumulative returns of the Fama-French five factors and momentum factor over the previous 60 months (*Return60*, *SMB60*, *HML60*, *RMW60*, *CMA60*, and *UMD60*).

	mean	std	min	25%	50%	75%	max
ENT	−0.0073	0.0613	−0.1404	−0.0499	−0.0135	0.0322	0.1881
ENT_NEWS	0.1542	0.0795	−0.0310	0.0965	0.1384	0.2023	0.4708
ENT_MODEL	−0.1615	0.0413	−0.3006	−0.1830	−0.1563	−0.1343	−0.0733
EPU	120.1456	45.0389	57.2026	88.2728	109.7164	144.1681	350.4598
SEN	0.0225	0.1964	−0.6361	−0.1022	0.0377	0.1775	0.4019
BEX	0.4059	0.2150	0.0878	0.2820	0.3624	0.4875	1.6233
JLN	0.6704	0.1245	0.5299	0.5924	0.6310	0.6920	1.2166
PCI	116.1411	38.9802	34.7400	84.6175	107.6350	143.0900	271.2900
VIX2	492.9679	474.4772	102.5248	209.2536	375.2209	600.2726	3927.3970
DGS10	3.4407	1.4030	0.5600	2.3500	3.4405	4.4775	6.6200
DGS10-2	1.1069	0.8725	−0.7200	0.2725	1.1010	1.8200	2.8700
DY	1.8094	0.3790	1.1100	1.6100	1.8200	2.0000	3.6000
CAY	−0.0072	0.0188	−0.0444	−0.0232	−0.0049	0.0073	0.0227
1/CAPE	−0.1132	0.1674	−0.5346	−0.2539	−0.0567	0.0366	0.0659
Return1	0.6438	4.6863	−17.2300	−2.0200	1.2300	3.4800	13.6500
Return12	8.0261	17.0219	−42.8302	0.0748	9.8582	18.0511	59.6603
Return60	51.4848	52.2836	−36.2647	1.3241	55.0996	95.4152	178.0090
SMB60	9.4205	25.0681	−50.7400	−7.8025	7.7350	20.6425	78.2400
HML60	6.6328	27.9585	−53.4300	−12.5600	2.2200	26.3650	91.0700
RMW60	19.8381	15.8996	−25.4300	7.1425	20.2550	30.6275	80.4300
CMA60	12.8265	22.5871	−19.2300	−2.5200	7.1000	21.6575	76.0000
UMD60	24.4751	34.9969	−49.7600	2.1925	18.9050	48.9200	108.9900

Table 2: Illustrative examples of high entropy articles. The table shows articles from three distinct months (first column) when *ENT* were high: January 2007, November 2010, and October 2014. Within each month, articles are ranked from the highest to lowest based on their raw entropy scores (last column). The displayed headlines (second column) and first sentences (third column) are from selected articles, whose entropy scores are within the top 1% for their respective month.

Month	Headline	Body (first sentence)	Score
2007-01	Freddie sells more than half $7 bln notes overseas	Freddie Mac says it sold 58% of its $4 billion 2-year notes to overseas investors, with central bankers taking 46% of the offering.	6.200
2007-01	US House sends business issues to Senate for action	Raising the minimum wage, changing Medicare drug purchasing and other initiatives important to the business community are landing on the U.S. Senate's doorstep, having won rapid passage since Jan. 9 in the House of Representatives.	6.016
2007-01	UBS sees narrow GSE debt spreads through 2007	Agency and mortgage securities yield spreads will stay narrow through most of this year, UBS strategists Laurie Goodman and Ivan Hrazdira said in an investor conference call.	5.936
2007-01	Fed's Poole says lag in GSE reform leaves crisis risk	Lack of progress in reining in mortgage lenders Fannie Mae and Freddie Mac leaves the economy at risk of possible financial crisis, St. Louis Federal Reserve Bank President William Poole said on Wednesday.	5.754
2010-11	U.S. as Currency Manipulator? It's a Bit Rich	Breakingviews U.S. Editor Rob Cox says the Fed's QE policy fits its mandate, despite China's criticism over what it called an "indirect currency manipulation."	6.384
2010-11	Ireland Makes Concessions to Restore Confidence	Following S&P's downgrade of Ireland's credit rating to single A, the Ireland government said it will reduce current spending, cut minimum wage and maintain its debated corporate tax at 12.5 percent.	6.266
2010-11	QE2 Critics Put Easing on Agenda at G20 Meeting in Seoul	Emerging market economies critical of the Fed's asset-purchase program have said quantitative easing threatens to flood their economies with excess liquidity.	6.191
2010-11	Euro Slumps on Irish Debt Fears	The euro falls to a 5-week low on growing concerns Ireland will be forced to default on its debts.	5.912
2014-10	Wall St declines after Fed ends bond-buying program	U.S. stocks fell on Wednesday, adding to their earlier declines after the Federal Reserve ended its monthly bond purchase program, as had been expected.	4.524
2014-10	Futures lower, investors look ahead to GDP data	U.S. stock index futures were slightly lower on Thursday as investors looked ahead to a report on economic growth and continued to digest recent comments from the Federal Reserve.	4.187
2014-10	Wall St flat after GDP, but Visa lifts Dow	U.S. stocks were mostly flat on Thursday, as a strong read on third-quarter economic growth raised new questions about monetary policy, though strong results at Visa single-handedly put the Dow in positive territory.	4.116

Table 3: In-sample predictions of 12-month ahead cumulative market returns. Returns are measured in percent. *Return60*, *SMB60*, ..., *UMD60* convert factor returns to state variables as explained in Section 2. The pre-COVID period ends in 2019, i.e., the year-ahead returns on the left-hand side of the regression in Equation (6) do not extend past 2019. The coefficient estimates have been normalized by the standard deviation of the right-hand side variables. Robust t-statistics are in parentheses and are based on Newey–West standard errors with four lags.

	12-Month Ahead Cumulative Return					
	Full sample			Pre-COVID		
	(1)	(2)	(3)	(4)	(5)	(6)
ENT	-2.821^{***}		-2.607^{***}	-2.728^{***}		-2.694^{***}
	(-3.358)		(-3.055)	(-2.999)		(-2.973)
EPU		3.841^{***}	3.436^{***}		1.239	0.937
		(3.185)	(2.889)		(0.946)	(0.750)
SEN	0.399	2.518	1.744	3.274^{*}	4.411^{**}	3.484^{*}
	(0.231)	(1.479)	(1.032)	(1.782)	(2.365)	(1.868)
VIX2	3.856^{***}	2.217^{*}	2.699^{**}	1.613	0.712	1.376
	(2.703)	(1.696)	(2.011)	(1.246)	(0.529)	(1.051)
DGS10	-1.204	-1.550	-0.557	-0.480	-1.304	-0.281
	(-0.927)	(-1.101)	(-0.423)	(-0.406)	(-0.984)	(-0.236)
DGS10-2	-4.269^{***}	-4.762^{***}	-3.682^{***}	-1.722	-2.839^{*}	-1.647
	(-3.139)	(-3.299)	(-2.709)	(-1.194)	(-1.883)	(-1.131)
DY	16.164^{***}	16.153^{***}	16.894^{***}	16.679^{***}	16.462^{***}	16.906^{***}
	(8.529)	(8.288)	(8.984)	(10.353)	(9.268)	(9.981)
CAY	-6.445^{***}	-4.317^{**}	-5.803^{***}	-3.963	-2.283	-3.950
	(-3.319)	(-1.994)	(-2.936)	(-1.486)	(-0.780)	(-1.473)
1/CAPE	-3.402	-2.516	-3.607^{*}	-2.462	-1.639	-2.601
	(-1.625)	(-1.095)	(-1.734)	(-1.194)	(-0.703)	(-1.244)
Return1	3.168^{***}	3.164^{***}	3.098^{***}	2.396^{***}	2.415^{***}	2.441^{***}
	(5.700)	(5.633)	(5.507)	(3.721)	(3.735)	(3.756)
Return12	9.248^{***}	8.939^{***}	9.046^{***}	8.242^{***}	8.063^{***}	8.220^{***}
	(6.554)	(6.220)	(6.436)	(6.181)	(6.007)	(6.149)
Return60	-4.773^{*}	-4.365	-3.460	-4.381	-5.662	-3.911
	(-1.691)	(-1.552)	(-1.214)	(-1.107)	(-1.421)	(-0.989)
SMB60	-14.645^{***}	-15.086^{***}	-14.077^{***}	-12.970^{***}	-14.321^{***}	-12.898^{***}
	(-5.406)	(-5.738)	(-5.267)	(-5.090)	(-5.646)	(-5.084)
HML60	-7.100^{***}	-5.980^{**}	-5.795^{**}	-5.927^{**}	-5.986^{**}	-5.650^{**}
	(-2.688)	(-2.216)	(-2.229)	(-2.090)	(-2.088)	(-2.001)
RMW60	4.574^{**}	4.101^{*}	4.779^{**}	6.382^{***}	5.579^{**}	6.396^{***}
	(2.014)	(1.841)	(2.172)	(2.849)	(2.398)	(2.873)
CMA60	19.474^{***}	19.598^{***}	18.537^{***}	14.566^{***}	15.994^{***}	14.562^{***}
	(5.472)	(5.848)	(5.329)	(4.040)	(4.480)	(4.027)
UMD60	-0.606	0.139	-0.093	0.112	0.346	0.217
	(-0.388)	(0.090)	(-0.061)	(0.074)	(0.218)	(0.144)
Note:						$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table 4: This table shows out-of-sample R-squareds, calculated using (8), of 12-month ahead market return forecasts using monthly rolling estimates of (7) in X-month windows. The different Xs correspond to the columns of the table. The methodology is explained further in Section 4.2. The data series are explained in Section 2.

Out-of-Sample R-Squareds

	Prediction Window in Months				
	12	15	18	21	24
ENT	−0.019	0.024	0.047	0.051	0.046
ENT_NEWS	0.069	0.087	0.096	0.091	0.060
ENT_MODEL	−0.074	−0.061	−0.059	−0.057	−0.084
EPU	−0.272	−0.322	−0.414	−0.417	−0.478
SEN	−0.595	−0.642	−0.678	−0.668	−0.607
VIX2	−1.320	−1.626	−1.813	−2.130	−2.422
DGS10	−0.214	−0.221	−0.284	−0.296	−0.312
DGS10-2	−0.246	−0.400	−0.511	−0.605	−0.637
DY	−2.001	−1.982	−1.722	−1.580	−1.767
CAY	−0.123	−0.141	−0.137	−0.136	−0.138
1/CAPE	−1.597	−1.964	−2.227	−2.269	−2.170
Return1	−0.059	−0.081	−0.075	−0.050	−0.060
Return12	−1.805	−1.691	−1.157	−0.855	−0.749
Return60	0.038	0.001	0.075	0.087	0.027
SMB60	−0.427	−0.532	−0.568	−0.573	−0.584
HML60	−0.799	−0.797	−0.826	−0.935	−1.082
RMW60	−0.170	−0.124	0.020	0.054	0.049
CMA60	−0.913	−0.977	−1.023	−1.143	−1.232
UMD60	−0.732	−0.705	−0.546	−0.481	−0.460
BEX	−1.206	−1.326	−1.715	−2.034	−2.237
JLN	−2.037	−1.999	−1.887	−1.754	−1.493
PCI	−0.133	−0.125	−0.158	−0.163	−0.216

Table 5: This table presents summary statistics of variables at the daily frequency used in this study, including Fama-French five factors market return minus risk-free rate (MKT), small minus big (SMB), high minus low (HML), robust minus weak (RMW), conservative minus aggressive (CMA), and momentum factor (UMD). $F_{EPU\_Momentum}$ represents the economic policy uncertainty factor mimicking portfolio using size-momentum returns as base asset. F_{ENT} , F_{ENT_NEWS} , F_{ENT_MODEL} represent entropy, news updates, and model updates mimicking portfolios. We also use size-book to market, size-investment, and size-profitability returns as base assets.

	mean	std	min	25%	50%	75%	max
MKT	0.0294	1.2682	−12.0000	−0.5100	0.0700	0.6300	11.3500
SMB	0.0090	0.6425	−4.5500	−0.3500	0.0200	0.3800	5.7100
HML	0.0062	0.7828	−5.0000	−0.3400	−0.0100	0.3200	6.7400
RMW	0.0173	0.5487	−3.0100	−0.2500	0.0100	0.2800	4.5200
CMA	0.0136	0.4606	−5.8700	−0.2100	0.0000	0.2200	2.5300
UMD	0.0171	1.0637	−14.3700	−0.4200	0.0700	0.5200	7.1200
$F_{EPU\_Momentum}$	−0.0954	2.5804	−22.3755	−1.4273	−0.1415	1.2136	20.5997
$F_{ENT\_Momentum}$	−0.0001	0.0035	−0.0246	−0.0018	−0.0000	0.0016	0.0367
$F_{ENT\_NEWS\_Momentum}$	−0.0001	0.0029	−0.0218	−0.0015	−0.0001	0.0014	0.0325
$F_{ENT\_MODEL\_Momentum}$	−0.0000	0.0018	−0.0129	−0.0009	0.0000	0.0010	0.0102
F_{EPU_BM}	−0.0547	2.0073	−13.2736	−1.0804	−0.0717	0.9715	16.7301
F_{ENT_BM}	−0.0001	0.0027	−0.0183	−0.0016	−0.0001	0.0014	0.0218
$F_{ENT\_NEWS\_BM}$	−0.0001	0.0028	−0.0164	−0.0016	−0.0001	0.0015	0.0187
$F_{ENT\_MODEL\_BM}$	−0.0000	0.0017	−0.0120	−0.0010	0.0000	0.0009	0.0100
F_{EPU_INV}	−0.0207	2.2522	−18.3677	−1.2390	−0.0525	1.1925	18.1130
F_{ENT_INV}	−0.0001	0.0029	−0.0287	−0.0017	−0.0002	0.0014	0.0199
$F_{ENT\_NEWS\_INV}$	−0.0001	0.0034	−0.0290	−0.0020	−0.0002	0.0017	0.0276
$F_{ENT\_MODEL\_INV}$	0.0000	0.0019	−0.0114	−0.0011	−0.0000	0.0011	0.0093
F_{EPU_OP}	−0.0039	1.9306	−15.5703	−0.9966	−0.0204	0.9854	14.9994
F_{ENT_OP}	−0.0001	0.0032	−0.0236	−0.0018	−0.0001	0.0016	0.0181
$F_{ENT\_NEWS\_OP}$	−0.0001	0.0032	−0.0232	−0.0018	−0.0001	0.0016	0.0207
$F_{ENT\_MODEL\_OP}$	−0.0000	0.0015	−0.0100	−0.0009	−0.0000	0.0008	0.0087

Table 6: Fama-MacBeth regressions using *MKT* (market minus risk-free rate), *SMB* (small minus big), *HML* (high minus low), *RMW* (robust minus weak), *CMA* (conservative minus aggressive), *UMD* (momentum factor), all of which are obtained from French's website, along with factor mimicking portfolios with base assets corresponding to the test assets. The first four columns represent different base assets: size–momentum, size–book/market, size–investment, and size–profitability. The last column represents the sign of the corresponding coefficient in Table 3 for forecasting 12-month ahead market returns if that coefficient is significant at 5% level. Risk premia coefficients are in percent per month. The risk premia coefficients are scaled by the standard deviations of betas in the first stage of Fama-MacBeth. General method of moments t-statistics are in parentheses.

Fama-MacBeth Factor Risk Premia (scaled coefficients)

	Base Assets the Same as Test Assets: Size -				
	Momentum	Book/Market	Investment	Profitability	Mkt-RF
	(1)	(2)	(3)	(4)	(5)
MKT	0.048^{*}	0.083^{**}	0.064^{**}	0.056^{**}	
	(1.681)	(2.189)	(2.023)	(1.997)	
SMB	0.100	0.067	0.008	0.079	−
	(1.371)	(0.907)	(0.108)	(1.051)	
HML	0.078^{*}	0.048	0.136^{***}	0.041	−
	(1.754)	(0.601)	(2.879)	(0.905)	
RMW	−0.066	0.062	−0.018	0.119^{*}	+
	(−1.050)	(1.283)	(−0.337)	(1.953)	
CMA	0.078^{**}	−0.007	0.065	0.018	+
	(2.142)	(−0.166)	(1.498)	(0.438)	
UMD	0.149	0.083^{***}	0.064^{**}	0.029	
	(1.342)	(2.896)	(2.346)	(0.928)	
F_{EPU}	-0.063^{*}	-0.083^{**}	−0.008	−0.013	+
	(−1.746)	(−2.342)	(−0.326)	(−0.432)	
F_{ENT}	-0.060^{*}	-0.066^{**}	-0.091^{***}	-0.082^{**}	−
	(−1.757)	(−2.398)	(−2.669)	(−2.175)	

Note: *p<0.1; **p<0.05; ***p<0.01

Table 7: Fama-MacBeth regressions using *MKT* (market minus risk-free rate), *SMB* (small minus big), *HML* (high minus low), *RMW* (robust minus weak), *CMA* (conservative minus aggressive), *UMD* (momentum factor), all of which are obtained from French's website, along with factor mimicking portfolios with base assets corresponding to the test assets. The first four columns represent different base assets: size–momentum, size–book/market, size–investment, and size–profitability. The last column represents the sign of the corresponding coefficient in Table A2 for forecasting 12-month ahead market returns if that coefficient is significant at 5% level. Risk premia coefficients are in percent per month. The risk premia coefficients are scaled by the standard deviations of betas in the first stage of Fama-MacBeth. General method of moments t-statistics are in parentheses.

Fama-MacBeth Factor Risk Premia (scaled coefficients)

	Base Assets the Same as Test Assets: Size -				
	Momentum	Book/Market	Investment	Profitability	Mkt-RF
	(1)	(2)	(3)	(4)	(5)
MKT	0.050*	0.077**	0.053**	0.051**	
	(1.720)	(2.157)	(2.022)	(2.000)	
SMB	0.093	0.071	0.010	0.082	−
	(1.287)	(0.966)	(0.131)	(1.095)	
HML	0.081*	0.054	0.133***	0.035	
	(1.780)	(0.684)	(2.766)	(0.758)	
RMW	−0.065	0.067	−0.014	0.122**	
	(−1.038)	(1.316)	(−0.265)	(1.978)	
CMA	0.071**	−0.013	0.063	0.009	+
	(2.055)	(−0.326)	(1.457)	(0.216)	
UMD	0.148	0.086***	0.066**	0.030	
	(1.335)	(3.032)	(2.332)	(0.937)	
F_{EPU}	−0.063*	−0.083**	−0.013	−0.009	+
	(−1.764)	(−2.297)	(−0.484)	(−0.311)	
F_{ENT_NEWS}	−0.044	−0.060**	−0.055**	−0.059*	
	(−1.426)	(−2.237)	(−2.000)	(−1.771)	

Note: *p<0.1; **p<0.05; ***p<0.01

Table 8: In-sample predictions of 12-month ahead economic policy uncertainty (*EPU*) and fundamental variables including unemployment rate (*UNRATE*), year-over-year change of industrial production (*INDPRO_YOY*), year-over-year change of the consumer price index (*CPI_YOY*), interest rates (*DGS10*, *DGS2*, and *DGS10-2*), the Chicago Board Options Exchange's CBOE Volatility Index (*VIX*), and the S&P500 earnings per share (*EPS*). The columns correspond to different dependent variables in (16). Each specification includes the lagged value of the dependent variable in question and the lagged value of all other macro variables as controls. The lagged value of one entropy measure (*ENT*, *ENT_NEWS*, or *ENT_MODEL*) is also used as a control, corresponding to each row. The coefficient estimates have been normalized by the standard deviation of entropy, *ENT_NEWS*, or *ENT_MODEL* respectively. This table summarizes results in Online Appendix Tables A9, A10, and A11. Robust t-statistics are in parentheses and are based on Newey–West standard errors with four lags.

Macro Forecasting Using Entropy and Other Macro Controls

	12-Month Ahead Macro Variables								
	EPU	UNRATE	INDPRO_YOY	CPI_YOY	DGS10	DGS2	DGS10-2	VIX	EPS
ENT	11.936	0.723***	−1.348***	−0.371**	−0.261***	−0.371***	0.049	1.571***	−4.890**
	(1.608)	(4.050)	(−3.248)	(−2.183)	(−3.569)	(−3.552)	(0.807)	(3.360)	(−2.239)
ENT_NEWS	7.779	0.642***	−1.213***	−0.356*	−0.182**	−0.351***	0.065	2.059***	−5.242**
	(1.333)	(4.441)	(−2.809)	(−1.906)	(−2.509)	(−3.133)	(0.999)	(3.786)	(−2.068)
ENT_MODEL	3.892	−0.079	0.178	0.090	−0.063	0.080	−0.045	−1.397**	2.197
	(0.530)	(−0.486)	(0.441)	(0.606)	(−1.009)	(0.762)	(−0.791)	(−1.994)	(0.909)

Note: *p<0.1; **p<0.05; ***p<0.01

Table 9: In-sample predictions of 12-month ahead market returns, measured in percent, using the specification in (17). Uncertainty measures include Bekaert et al. (2022) (*BEX*), Jurado et al. (2015) (*JLN*), the Azzimonti (2018) Partisan Conflict Index (*PCI*), and squared implied volatility (*VIX2*). The other control variables are explained in Section 2. Full results are shown in Online Appendix Table A17. The coefficient estimates have been normalized by the standard deviation of the right-hand side variables. Unscaled full results are shown in Online Appendix Table A18. Robust t-statistics are in parentheses and are based on Newey–West standard errors with four lags.

In-Sample Return Forecasting with Uncertainty Controls

	12-Month Ahead Cumulative Return					
	(1)	(2)	(3)	(4)	(5)	(6)
ENT	-2.266^{**}	-2.530^{***}			-2.306^{**}	-2.413^{***}
	(-2.545)	(-3.040)			(-2.522)	(-2.709)
EPU			4.839^{***}	3.674^{**}	4.686^{***}	3.531^{**}
			(3.928)	(2.502)	(3.924)	(2.385)
SEN			2.091	2.319	1.319	1.821
			(1.212)	(1.465)	(0.768)	(1.158)
Other Uncertainty	No	Yes	No	Yes	No	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.648	0.671	0.657	0.667	0.666	0.677

Note: *p<0.1; **p<0.05; ***p<0.01