Economic Narratives and Market Outcomes: A Semi-supervised Topic Modeling Approach *

Dat Mai† Kuntara Pukthuanthong‡

December 2021

Abstract

We employ sLDA to extract the narratives discussed by Shiller (2019) from 7 million NYT articles over 150 years. The estimation addresses look-ahead bias and changes in semantics. Panic and the narrative index positively predict market return and negatively predict volatility. Panic presents time-varying risk aversion. The narrative predictability increases recently at both market and portfolio and monthly and daily intervals. The narrative index constructed from 2 million WSJ articles over 130 years retains its predictive power, but Stock Bubble emerges as a negative market predictor. Media customizes their narratives to their readers, having a diverse effect on the market.

*We are indebted to Richard Roll for inspiration and brilliant suggestions. We thank Vineer Bhansari, Murray Frank, John Howe, Tim Loughran, Michael O'Doherty, Clive Walker, Russ Wermers, and the seminar participants at the University of Missouri-Columbia, the University of Missouri-St. Louis, Missouri State University, the Doctoral Tutorial at the European Finance Association (2021), the PhD Student Symposium at the University of Texas at Austin (2021), the Chicago Quantitative Alliance (CQA) Annual Academic Competition (2021), the Northern Finance Association Meeting (2021), the Financial Management Association Meeting (2021), the Southern Finance Association Meeting (2021), the Financial Research Association Meeting (2021), and LongTail Alpha LLC for helpful comments. We acknowledge computing support from the Research Computing Support Services (RCSS) at the University of Missouri-Columbia.

†Graduate student, Department of Finance, University of Missouri-Columbia, Email: datmai@mail.missouri.edu

‡Professor of Finance, Department of Finance, University of Missouri-Columbia, Email: pukthuanthongk@missouri.edu

Conflict-of-interest disclosure statement

Dat Mai
I have nothing to disclose.

Kuntara Pukthuanthong
I have nothing to disclose.

"What people say about the economy can set off a recession."[1]
—Robert J. Shiller, *New York Times*, September 12$^{\text{th}}$, 2019.

1 Introduction

Shiller (2017, 2019) introduces narrative economics which states that popular stories weaved into daily conversation, such as how much to save and consume, as well as when and where to invest, can eventually affect individual and collective human behaviors and thus drive economic outcomes. According to Shiller, studying these economic narratives can enhance our ability to predict or prepare for future major economic events. Shiller (2019) points out nine major economic narratives that have been mutating over the past two centuries, namely, *panic versus confidence*, *frugality versus conspicuous consumption*, *gold standard versus bimetallism*, *unemployment due to labor-saving machines*, *automation and artificial intelligence replacing human jobs*, *real estate booms and busts*, *stock market bubbles*, *boycotts and evil business*, and *wage-price spiral and evil labor unions*. Although narrative economics is intuitive, simple to understand, and important, the theory has received little attention from most economists because of its lack of a well-developed framework. In this study, extracting narratives from seven million articles from the *New York Times* (NYT) over 150 years and two million articles from the *Wall Street Journal* (WSJ) over 130 years, we attempt to answer: Can time-varying changes in news narratives predict stock market outcomes? Does the source of narratives matter?

To answer the above two research questions, we first need to measure variations in the amount of attention paid to narratives over time, and we will use topic weights to capture variations. Shiller (2019) discusses how these market-moving stories are transmitted by word of mouth, news media, and, increasingly, social media. In this study, we capture narratives from news rather than other sources because the news contains both fundamentals and unobservables, including rumors, speculations, and expectations (Tetlock, 2007; Garcia, 2013). News is a plausible proxy for investors' beliefs because the press has demand-side incentives to cater its content to its readers' beliefs (Shiller, 2005). Confirmation bias, as it is understood in the psychology literature, refers to the phenomenon of individuals seeking out information consistent with their beliefs. The notion that readers seek out media sources in line with their own political beliefs gained public traction in the aftermath of the 2016 election. Mullainathan and Shleifer (2005)

[1] https://www.nytimes.com/2019/09/12/business/recession-fear-talk.html

summarize the communications, psychology, memory, and information processing literatures that support the notion that people receive utility from content consistent with their beliefs. Gentzkow and Shapiro (2010) provide empirical evidence in support of this theory showing that media slant is largely attributed to consumer preferences. As Shiller (2019) uses countless excerpts from the NYT as support for his arguments, we present the main results based on NYT[2] and robustness test results based on WSJ. We present economic insights from applying two different sources.

We employ a topic modeling approach called seeded latent Dirichlet allocation (Lu et al., 2011, henceforth sLDA) to extract Shiller's narratives. sLDA is a recent extension of the canonical unsupervised LDA model (Blei et al., 2003; Griffiths and Steyvers, 2004), which has experienced burgeoning popularity in computer science and other social science fields. Rather than letting the model freely cluster words based on their co-occurrences like in LDA, users of sLDA offer seed words to guide the convergence of topics toward their predefined themes.

In this paper, the application of sLDA allows us to address two issues: look-ahead bias and changes in semantics (i.e., how word usage evolves over time). A study of predictability needs to be free from look-ahead bias and any study of narratives using 150 years of data needs to control for a change in word usage. We reclassify the narratives in Shiller (2019) into 10 narratives based on their content similarity. As inputs into the sLDA model, key terms (or seed words) for each topic are selected based on Shiller (2019), yet we add certain words representative of the narratives to help the sLDA model better extract the narrative weights. We design a rolling estimation process to extract narratives from news articles using only news data over the past 10 years. This estimation scheme is especially important in out-of-sample tests as it avoids any look-ahead bias associated with using future news articles in extracting topic weights today. The output of the estimation process is the article-level vectors of weights, the elements of which represent the proportion of content (or attention) devoted to the corresponding narrative. From the article-level topic weights, we compute the daily and monthly time series of topic weights for each narrative. We then use these time series to predict the excess market return and market volatility. The idea of using seed words is to list the key words for each topic and let the model determine the relevant ones based on word co-occurrences by context (the past 10 years of news articles). This approach allows us to capture the

[2]The NYT is one of the most prestigious newspapers in the world and has been relied on in finance research (see, e.g., Garcia (2013) or Hillert and Ungeheuer (2019)). To date, the NYT has received 130 Pulitzer Prizes, almost double its nearest competitor. See https://www.nytco.com/company/prizes-awards/.

words popularly used during that time and consider how word usage evolves.

It is important to note that the use of the Shiller (2019)'s seed words does not induce look-ahead bias into the topic weights extracted in the periods before the existence of the terms. For example, the term "great depression" only emerges during the 1930s, after the Great Depression. In fact, the sLDA model is designed to exactly address this situation. In the sLDA model's setup, if a word does not show up, that is, has zero count, in the set of documents, it does not enter the posterior topic-word distribution and hence has no influence on the output. Although some words might seem to be increasingly relevant recently, our topic model focuses on their relative frequency and thus these topic weights are much lower in the last two decades than in their peak in the mid-twentieth century. For example, financially adjacent terms, such as "great recession" in Panic, "bitcoin, cryptocurrency" in Monetary Standard, or "artificial intelligence, internet, machine learning" in Technology Replacing Jobs, have been only recently coined and popularized. Our model does not put higher weights to these topics in recent periods than in the distant past. Figure 4 and Figure B2 show the topics Panic, Money, or Tech do not present an upward trend in the time series of their weights. Thus, we want to emphasize, that, even though the topics were introduced recently, our topic model focuses on their relative frequencies and the frequencies of words co-occurring with them.

Our sample covers a long time series of narratives: over 150 years from NYT and 130 years from WSJ, both since their inceptions. Using a long time series from two different sources is important for four reasons. First, our data of 150 years cover many rounds of business cycles and rare disaster events. Asset pricing literature shows how the expectation changes around rare disasters and influences equity risk premium (see for instance, Rietz (1988), Barro (2006), and Julliard and Ghosh (2012)). Our sample includes events such as wars, pandemics, depressions, natural disasters, and terrorists. History tends to repeat itself (Chava et al., 2020); thus, researchers use the past to learn how the market responds to rare disaster risks. This can be seen from the recent pandemics. Roberts and Tehrani (2020) show public health response to disease outbreaks has remained nearly unchanged between the 2019 coronavirus and the 1918 influenza pandemics. Berkessel et al. (2021) show pandemics initially spread among people of higher social status (not lower) for both pandemics. Our long time series data allows us to study the effects of real-world events on the predictability of the narratives and whether the predictability of the narrative indexes is cyclical and time varying that shorter time series cannot.[3]

[3]These results are presented in Section 6.3 and Section 7.3.

Second, it provides us an "out-of-sample" test whether the results in the recent US period (we show the predictability of the narratives is strongest after 2000s) are generalizable across times and sources (Schwert, 1990). Using the extended time series from two different national media, we can increase the statistical power to reject the null of no return predictability.

Third, our sample combats two potential biases that affect inferences about expected performance and risk. Survivorship bias, conditioning on eventual market outcomes and return data series with no disruption, produces an upward bias in performance relative to ex ante expectation (Brown et al. (1995)). We mitigate concerns over this bias by studying articles from NYT and WSJ since the newspapers inception, thereby allowing us to characterize the whole distribution and bypass any changes that determine the survival of the media. Another bias is incurred from easy accessibility and availability of the data (Dimson et al. (2009)). Our sample alleviates this bias by covering both NYT and WSJ articles from all newspaper sections. Fourth, using 150 years is consistent with Shiller (2019) who suggests nine major economic narratives that mutate over the past two centuries. To the best of our knowledge, we are the first paper that analyzes news articles from all newspaper sections of NYT and WSJ since their inception. The material and computation costs to analyze nearly nine million articles are prohibitive and almost insurmountable.

Among the narratives extracted from the NYT, the most important one is Panic. Although Panic as discussed by Shiller (2019) centers on financial panic and economic recessions, we widen the theme to encompass wars, tension, and epidemics by injecting these terms as seed words into the model. The literature on rare disaster risks (Barro, 2006, 2009; Gabaix, 2012), which incorporates the probability of rare disasters, mostly wars, into explaining the high equity premium, inspires this expansion. The most common terms of Panic as output from the sLDA model are *panic, fear, crisis, depression, hard time, war, america, state, government, tension,* and *epidemic.*[4] In tracking the articles making the biggest contribution to the Panic index over the past 30 years, we find that all of them feature wars, terrorism, and tension in international relations.

Regarding the impact of Panic on stock market outcomes, we find that Panic can strongly positively predict market returns and negatively predict market volatility, both

[4]Terms such as *state* and *government* by themselves are obviously not panic-related words. However, in the context of other words showing up with them in Panic, these terms are an integral part of a topic on international tensions. In other words, we cannot talk about international tensions without mentioning *state* and *government*. Hence, in topic modeling, it is more about examining the thematic content of a topic as a whole, and less about picking up on each individual words belonging to the topic.

realized and implied. The predicting power of Panic over market risk premium increases over time. For example, over the whole sample of 150 years, a one-standard-deviation increase in Panic predicts a 3.44% increase in annualized excess returns in the next month and the monthly in-sample R^2 is 0.31%, while over the past 20 years, the respective numbers are 10.85% and 4.22%. Panic is significant for both subperiods (1871–1949 and 1950–2019).

In addition to Panic, we construct a narrative index from 10 narratives via the two-step partial least squares (PLS) approach of Kelly and Pruitt (2013, 2015). The PLS index heavily loads on Panic with a correlation of 81%. We thus interpret the PLS index as a stronger version of Panic, which inherits all of its predicting features. Indeed, we find that the monthly predictive regression of market returns with the PLS index yields a slope of 5.6% and an R^2 of 0.91% over the whole sample and a slope of 11.75% and an R^2 of 5.02% over the past 20 years. For the subperiods (1871–1949, 1950–2019), the PLS index remains a significant predictor at at least the 5% level. The predicting power of the PLS index over market returns is not subsumed by common macroeconomic variables, such as those in Goyal and Welch (2008), the output gap in Cooper and Priestley (2009), or short interest in Rapach et al. (2016); other uncertainty indexes, such as the implied volatility (VIX), news-implied volatility (NVIX) in Manela and Moreira (2017), financial and macro uncertainty in Jurado et al. (2015), economic policy uncertainty in Baker et al. (2016), or disagreement index in Huang et al. (2020); other sentiment variables, such as news sentiment, investor sentiment in Baker and Wurgler (2006) and Huang et al. (2015), or manager sentiment Jiang et al. (2019); and confidence indexes constructed by Professor Shiller, such as one-year confidence index and crash confidence index.

From a theoretical viewpoint, we find that the predictability of Panic (and thus the PLS index) is consistent with the intertemporal capital asset pricing model (ICAPM) of Merton (1973) when Panic proxies for time-varying relative risk aversion (RRA). Modeling the unobserved RRA as a linear function of Panic and estimating the univariate ICAPM model via the GARCH-M framework as in Lundblad (2007), we find that Panic enhances the risk-return trade-off across different GARCH-M specifications and time samples. Specifically, the loading on Panic is significantly positive, and the GARCH specification with time-varying RRA featuring Panic produces a substantially higher R^2 than the one without Panic. These results indicate that risk aversion and Panic are positively correlated. The univariate ICAPM also implies that when investors become more risk averse, they require a higher market risk premium and/or lower market volatility to hold the market portfolio. As risk aversion is reasonably assumed to rise when Panic is high, Panic is then expected to positively predict market return and negatively predict market volatility, an implication confirmed by our empirical results.

To disentangle the predictability of narratives reported by the news media from actual events, we rerun the predictive regressions with Panic and control for real-world event indicators. Specifically, we create event dummy variables equal to one if there is at least one such event in a month and zero otherwise for the following events: recession, bank failure, war, disaster, epidemic, and a combination of all events. When both Panic and each event indicator are included in the monthly return prediction, the prediction of Panic remains intact, while only recession significantly negatively predicts the market. This finding rules out the alternative explanation that Panic proxies for changes in time-varying RRA triggered by the real-time stressful events.

To investigate whether economic narratives create economic value for real-time investors, we conduct standard out-of-sample tests in the predictability literature. With the expanding window estimation, Panic outperforms all individual predictors studied in this paper in terms of out-of-sample R^2 (R^2_{OS}), which compares the forecasting power of a predictor against the historical mean return used as a forecast. A positive R^2_{OS} indicates the predictor outperforms the historical mean. Panic via a standard ordinary least squares (OLS) regression yields an R^2_{OS} of 0.28% over the whole evaluation window and 1.41% since 2000. Combining the forecasting power of all narratives via PLS produces an R^2_{OS} of 0.24% and 1.71% over the whole sample and the last two decades, respectively. We conjecture the stronger predictive power after 2000 is driven by the digitalization of news and the fast speed of information diffusion. Our results support the evidence shown by Obaid and Pukthuanthong (2021), who document sentiment in photos and text from news has a strong market predictability after 2000.

In terms of asset allocation implications, we consider a mean-variance investor who allocates his portfolio between the stocks and a risk-free asset using either the return predictive model or the historical mean return to guide the portfolio weights. With a risk aversion coefficient of three, we find the economic gains for the investor utilizing narratives in forming his portfolios increase over time, consistent with the R^2_{OS} results. Interestingly, using random forest—an advanced machine learning method—and all narratives to predict future returns yields an annualized utility gain of 4.45% over the past 20 years. Using Panic alone or a combination of narratives to guide our portfolio decisions allows us to achieve a higher Sharpe ratio than that achieved with a simple buy-and-hold strategy.

To validate the predictive power of the narrative index, we apply sLDA to extract narratives from two million WSJ articles over the past 130 years. We find the narrative index remains a strong predictor of market returns. Interestingly, we find Stock Market Bubble carries the most weight and is a negative predictor of market returns over the past 20 years with an average R^2_{OS} of 2.72% relative to 1.12% from Panic in the NYT. We conjecture that the target audiences for WSJ and NYT explain the difference in the newspapers' most influential topic. WSJ targets financial market participants, whereas NYT targets a well-educated audience with broad interests.

To further explore the predictive power of Panic and the narrative index, we extend the analysis to the characteristics portfolios sorted on industry, size, value, and past returns. We document that Panic and the narrative index continue to positively predict returns on these portfolios, although the degrees of exposure to narratives differ across assets. Furthermore, the predictability of Panic and the PLS index over the stock market returns continues to hold at the daily frequency.

We conduct a robustness check to investigate whether the sLDA model yields any additional economic insights beyond a simple count of topic-based seed words in the news archive, that is, the traditional dictionary-based approach. To answer this question, we simply compute topic weights based on seed word frequencies in the NYT articles. We find that with these manually counted topic weights, the predictability of narratives is subsumed by other economic and uncertainty variables in the monthly return prediction. Moreover, in an out-of-sample analysis, the manually counted topic weights produce much lower R^2_{OS}'s than the statistically constructed ones.[5] Hence, these results indicate that the dictionary-based approach misses important content for each narrative, and sophisticated statistical methods prevail by recognizing and accounting for the missing content. Frankel et al. (2021) show measures based on machine learning offer a significant improvement in explanatory power over dictionary-based measures.

To our knowledge, the paper most related to this study is Bybee et al. (2021). However, we can point to three main differences between the two papers. First, the main research question asked in Bybee et al. (2021) is how to use news content to reconstruct various financial and macroeconomic variables, while in this paper, we ask whether the specific narratives in Shiller (2019) can predict future stock market outcomes. In other words, the former is about contemporaneous regressions, while the latter is about predictive regressions. The first main difference leads to the second difference in the statistical topic models used in extracting contents from news. Bybee et al. (2021) employ the traditional unsupervised LDA model and use cross-validation to select 180 topics in the

[5]To conserve space, we report these results in Internet Appendix D.

model. They require a large number of topics so that different topics can optimally match different economic variables, stemming from their goal to reconstruct those variables from news. In contrast, we employ a semisupervised LDA model in which we inject initial seed words to extract the desired themes. Our model uses only 11 topics (10 seeded, plus one unseeded, topics) as we need to extract only the specific topics of interest and the intent is not to discover all thematic contents from news. The sLDA model is also better suited to the predictive analysis as it allows for consistent topic contents via rolling estimation of the model to avoid look-ahead bias and incorporate changes in semantics, two features infeasible under the unsupervised LDA model. While Bybee et al. (2021) look at the business- and finance-related articles in the WSJ over the past 30 years, we examine all articles in the NYT and WSJ over the past 150 years.

This paper contributes to different branches of the literature. First, and foremost, our paper is related to the newly proposed theory of narrative economics. Specifically, this study is among the first to test the ideas in Shiller (2019) on a large scale using nearly seven million NYT news articles and two million WSJ news articles across all newspaper sections. Using a long time series (i.e., since the newspapers' inception) and two sources with differing political views and target audiences (both NYT and WSJ), which helps eliminate selection bias, we find the narrative index positively predicts stock market returns. Our sample combats both data bias and survivorship bias. Among the topics suggested by Shiller (2019), Panic, the most influential topic extracted from NYT, positively predicts market returns and negatively predicts volatility. Under the ICAPM framework, panic captures risk aversion, which tends to rise during panic times. On the other hand, Stock Market Bubble, the most influential topic extracted from WSJ, negatively predicts stock returns. As we observe a reversal, Stock Market Bubble seems to present mispricing. We leave it for future research to perform rigorous investigation as this is not the main goal of this paper. The most influential narrative from the source that targets a general audience that is educated and liberal like NYT captures risk aversion, whereas that extracted from the media targeting financial market participants who are conservative like WSJ tends to capture mispricing. This highlights that narratives depend on the sources and audience they target.

We want to stress an important distinction between the ideas of Shiller (2019) and our empirical results. While Shiller argues for causality between economic narratives and economic outcomes, we do not attempt to establish any causal relationship in our paper. While we hypothesize that risk aversion rises during stressful times, as captured by the attention paid to Panic in news, we do not argue that stress- and anxiety-related news cause risk aversion to increase, nor do we argue that stock-market-bubble-related news cause mispricing. Instead, stress- and anxiety-related news and stock-market-bubble-related news may reflect the levels of risk aversion and mispricing, respectively, among investors.

Second, this paper contributes to the literature on rare disaster risks, which incorporates disaster probabilities and disaster loss into the standard consumption-based model to explain the high equity premium (Barro, 2006, 2009; Gabaix, 2012; Wachter, 2013). Specifically, according to Barro (2009), an increase in the disaster probability leads to a rise in the equity premium. The Panic index consists of various themes related to wars, international tension, and epidemic. One could reasonably expect that the implied disaster probability increases when Panic is high, leading to an increase in the equity premium. This is consistent with our result that Panic is a strong predictor of future excess market returns.

This paper further relates to the recent literature on measuring political risks extracted from textual data (see, e.g., Baker et al. (2016); Hassan et al. (2019)). While these papers focus on firm-level responses to changes in political uncertainty, we direct our analysis to aggregate stock market outcomes. Moreover, while a number of studies, such as Pástor and Veronesi (2013) and Brogaard and Detzel (2015), document that the economic policy uncertainty (EPU) index in Baker et al. (2016) can positively predict the aggregate market return over the long term only, our Panic from NYT, Stock Market Bubble from WSJ, and narrative index are stronger predictors of short-term market returns.

Finally, this paper contributes to the burgeoning literature on applications of modern natural language processing tools in business and financial research. Digitized texts offer a rich and multidimensional resource with novel and unique insights into many economic relationships not captured by traditional economic data. An increasing number of papers have utilized advanced topic modeling tools to extract thematic contents from texts, such as Dyer et al. (2017), Choudhury et al. (2019), Brown et al. (2020), and Bybee et al. (2021). Unlike the majority of finance papers that use the traditional unsupervised LDA model, we employ a semisupervised LDA model, which suitably serves the purpose of extracting a predefined set of narratives from news. Indeed, recent papers in natural language processing, such as Lu et al. (2011), Jagarlamudi et al. (2012), Eshima et al. (2020), and Watanabe and Zhou (2020), have documented the advantages of a (semi)supervised LDA model over the unsupervised one. Among other preferable features, a guided LDA model ensures the post-estimation topic content is consistent with a priori expectations and avoids the post hoc labeling of topics, a common practice under unsupervised LDA.

2 Method

2.1 Model

In this paper, we employ an sLDA model (Lu et al., 2011) to extract news narratives. We first briefly describe the model. This discussion closely follows from Lu et al. (2011). Under the standard unsupervised LDA model of Griffiths and Steyvers (2004), a document, d, is generated under the following hierarchical process:

- For each topic k,
  - choose a topic-word distribution: $\phi_k \sim \text{Dirichlet}(\beta)$.
- For each document d,
  - choose a document-topic distribution: $\theta_d \sim \text{Dirichlet}(\alpha)$.
  - For each word w in document d,
    * choose a topic: $z_{d,w} \sim \text{Mulinomial}(\theta_d)$,
    * choose word: $w \sim \text{Multinomial}(\phi_{z_{d,w}})$.

To inject seed words (prior knowledge) into the model, Lu et al. (2011) specify a combined conjugate prior for each seed word, w, in $\phi \sim \text{Dirichlet}\left(\beta + C_w\right)_{w \in V}$, where C_w is a pseudo-count added to the topic to which w belongs. Taking these steps creates an asymmetric prior. When we have no prior knowledge for a word, w, $C_w = 0$. The LDA model can be estimated using Gibbs sampling from posterior distribution (for details, see Griffiths and Steyvers (2004)). With a sample obtained via Gibbs sampling, we can approximate the topic-word distribution, ϕ_k, for each topic k and the document-topic distribution, θ_d, for each document d.

2.2 Seed Words

The foundational piece of an sLDA model is the set of seed words representing the prior knowledge of each topic. Watanabe and Zhou (2020) emphasize that a dictionary of seed words needs to be carefully chosen based on field-specific knowledge independent of word frequencies in the collection of texts used. For that reason, the setup of a sLDA model is perfectly aligned with the task at hand, namely, extracting predefined narratives from news articles. Table 1 lists the lemmatized seed words for each narrative, all of which were manually collected from Shiller (2019).

As shown in Table 1, we have reclassified the 9 narratives from Shiller (2019) into 10 topics to facilitate our estimation. Specifically, as *panic* and *confidence* are opposing notions, we split them into two topics. Similarly, *frugality versus conspicuous consump-*

tion is split into frugality and conspicuous consumption. In contrast, we combine *labor saving machines* and *automation and artificial intelligence* into one topic because of their similarities. Among the topics, we have substantially expanded the scope of Panic to encompass various disaster themes, such as economic recessions, wars, international tensions, and epidemics.

In an unsupervised LDA model, a statistical criterion, such as posterior likelihood or perplexity score via cross-validation, can be used to find the optimal number of topics (see, e.g., Griffiths and Steyvers (2004)); however, an sLDA model lacks guidance or theory on how to pick the number of topics. Hence, the best approach is to examine the most common terms per topic post-estimation to determine whether the topics feature the desired contents. In addition to the 10 narratives from Shiller (2019), we include one additional "garbage collector" to absorb everything else in the news unrelated to these narratives. In unreported results, we find that increasing the number of unsupervised topics to five changes the main results only minimally.

2.3 Estimation

Figure 1 illustrates the rolling estimation scheme used in the paper. Specifically, at the end of each month t, we run the sLDA model using all news data over the past 120 months (months $(t-119)$ to t). We use 10 years of news data in the monthly estimation to balance the comprehensiveness of news data required to estimate the model and computational costs. On average, every 10 years of historical data consists of around 460,000 articles, which should be sufficient to reliably extract the topic weights at the time of estimation.[6] Notably, rolling estimation is viable only under the sLDA model because with the use of seed words, we can ensure the consistency of thematic content over time. In contrast, the unsupervised LDA model can generate inconsistent results and make controlling the topic weights and interpretability difficult. During each estimation, we draw 200 samples[7] from the posterior distribution of the sLDA model and use the last draw to estimate the document-topic weights θ_d; that is, we estimate a distinct 10×1 vector $\theta_d = [\theta_d^1, \theta_d^2, \ldots, \theta_d^{10}]$ for each news article, d, in the estimation window. We then compute the monthly weights

[6] Estimation is implemented by the seededla package in R and run on a high-performance computing (HPC) cluster. Full estimation of the model parallelized on 80 computational nodes requires about one week to complete. We keep the default values for hyperparameters α and β in the package.

[7] In addition to the number of topics and the number of articles, the number of sample draws from the posterior distribution is a computational cost consideration in any topic model. We believe 200 draws produce reliable estimate of the targeting posterior distribution, while further increases in the number of draws introduce unnecessary additional computational costs.

of each topic i ($i = 1, 2, \ldots, 10$) as the average weight of each topic across all articles in month t, weighted by the length of each article:[8]

$$\theta_t^i = \frac{\sum_{d=1}^{n_t} \theta_d^i \times length(d)}{\sum_{d=1}^{n_t} length(d)}, \tag{1}$$

where θ_t^i is the weight of topic i in month t, n_t is the total number of news articles in month t, and $length(d)$ is the number of ngrams in article d.[9]

Although 10 years of news articles are used to estimate the model, the final topic weights in month t are computed from the news articles of that month only. The final output of the estimation process is a time series of monthly weights for each of the 10 narratives. These time series will be used as an input into our economic forecasting application.

Our method takes the evolution of word usage into account. Although the list of Panic seed words remains unchanged, the model is re-estimated every month using data for the past 10 years (including the current month), so the actual words clustered in the Panic topic change monthly. That is, the list of unobserved Panic words (from the output of the model) changes month over month based on the change in language.

3 Data

We use two sets of data in this paper. The first set comprises news articles. Shiller (2019) discusses how market-moving stories are transmitted by word of mouth, news media, and, increasingly, social media. In this study, we investigate the news media channel, particularly NYT articles. The NYT, one of the most prestigious and circulated newspapers in the world, is an ideal laboratory to test the theory of narrative economics. The NYT cites its mission as a commitment to providing readers with a timely, objective, and comprehensive account of what is happening around the world. The time span of the NYT also aligns with the majority of narratives discussed by Shiller, and Shiller (2019) cites countless excerpts from the NYT as supporting arguments for his narratives. We perform a robustness check using articles from all newspaper sections in the WSJ since its inception in 1889. To streamline our presentation, we report the results of the WSJ in Section 8 and Internet Appendix C.

[8] Equal weighting of topic weights across articles yields similar results.

[9] An ngram is a sequence of n words. For instance, "San Diego" is a 2-gram, and "A study in narratives is needed" is a 6-gram.

Shiller (2019) stresses the importance of studying personal stories from all aspects of lives, such as personal letters, diaries, and even sermons. Hence, it is essential that we analyze all sections of the NYT; our approach is in stark contrast to Bybee et al. (2021), who limit their attention to articles appearing in sections directly related to economics and finance. However, we still remove articles with limited content, such as those that contain mostly numbers, names, or lists. Then we conduct the standard text processing steps. Internet Appendix A and Table B1 report the details. After the cleaning steps, for each month t, we create a document-term matrix containing all articles over the past 10 years up to and including the current month. Each row of the matrix is an article; each column is a term; and each entry is the count of that term in the article. The document-term matrix and topic-based seed words are input into the sLDA model to estimate monthly topic weights as described in the previous section.

Panel A of Figure 2 plots the time series of monthly article counts after the exclusion of articles with limited content. Since 1871, the NYT has published more than 6.8 million news articles with a monthly average of 3,800 articles.[10] Before the year 1900, the NYT published around 2,000 articles a month. The number of monthly articles increased gradually after 1900, hovering between 4,000 and 6,000 until the end of the twentieth century. Amidst industrywide struggles related to declining ad revenues and subscriber bases beginning in the 2000s, the NYT began scaling down their publishing capacity to around 2,000 articles a month during the 2010s.[11] However, the number of monthly articles surges back to just under 4,000 toward the end of the sample. A newspaper strike occurred from 1902 to 1903, and news articles spiked at the start of World War I.

Panel B of Figure 2 reports the average monthly article length, which is defined as the total count of unigrams (one-word terms), bigrams (two-word terms), and trigrams (three-word terms).[12]

While Bybee et al. (2021) consider only unigrams and bigrams in their paper, we extend the analysis to trigrams as a majority of the seed words have three words, such as *real estate boom, stock market bubble, and cost push inflation.* Over the whole period 1871–2019, articles come in at an average length of 493 ngrams. Articles tended to have around 500 ngrams until the 1920s. After that, they hovered just above 400 ngrams until the 1960s. Since then, article length has been on upward trend, reaching about 600 ngrams during the 2010s.

[10] Data are missing for September and October 1978 (due to strikes) and thus are excluded from Figure 2.

[11] For more details, see https://www.pewresearch.org/journalism/fact-sheet/newspapers/

[12] See Internet Appendix A for more details on the construction of ngrams.

The second set of data concerns stock market outcomes. The main analyses of this paper center on predictions of market returns and market volatility. We obtain the total S&P 500 index from Global Financial Data (GFD) with monthly data available from January 1871 and daily data available from January 1927.[13] We also require the risk-free rate, in order to compute excess returns. Daily and monthly risk-free rates are downloaded from Professor Kenneth French's website. For monthly risk-free rates before 1927, we use the series from Goyal and Welch (2008). We also conduct analyses with characteristics-sorted portfolios whose returns are downloaded from Professor Kenneth French's website. Finally, we obtain the implied volatility index (VIX), available from January 1990, from the Chicago Board Options Exchange (CBOE).

4 Economic Narratives

4.1 Contents of Economic Narratives

We are interested in the thematic content of each topic, that is, the most common words per topic. Our estimation approach is different from the traditional unsupervised LDA model in that we define a set of seed words per topic and re-estimate the model every month. Our estimation scheme is deemed successful when the extracted topics can uncover the predefined themes. Hence, to investigate the contents of the 10 extracted narratives, during every monthly estimation of the sLDA model, we retain the 30 most common ngrams per topic, that is, those having the highest $\phi_{z_{d,w}}$ in the sLDA model. Then the most important words for each topic over time are evaluated by the across-time frequencies of these words. To create visuals of each topic, we create word clouds using the top words from each topic. Across-time frequencies are used to determine the top words; the bigger the size, the higher the frequency of that word in the topic. To conserve space, we report the word clouds of five main topics (based on their weights in the PLS index discussed next) in Figure 3, and we report the remaining topics in Figure B1 in the Internet Appendix.

As indicated by Figure 3, the sLDA model performs well at extracting Shiller's narratives from the NYT articles. For example, the most common terms for Panic extracted

[13]The GFD description is as follows: "The S&P 500 Total Return Index is based upon GFD calculations of total returns before 1971… Beginning in 1871, data are available for stock dividends for the S&P Composite Index from the Cowles Commission and from S&P itself. We used this data to calculate total returns for the S&P Composite using the S&P Composite Price Index and dividend yields through 1970, official monthly numbers from 1971 to 1987, and official daily data from 1988 on."

by the model are [panic, fear, crisis, depression, recession, hard_time, epidemic, war, tension, government, american, united_state, etc.], all of which strongly overlap with the seed words. Accordingly, the extracted Panic index encompasses not only financial panic but also panic related to politics, wars, and diseases. The top words for Technology are [machine, invention, network, computer, unemployment, etc.]; for Real estate [real_estate, building, speculation, bust, crash, boom, bubble, etc.]; for Stock [stock, speculation, crash, boom, bubble, bust, bear, bull, margin, etc.]; and for Boycott [boycott, outrage, strike, moral, anger, community, protest, etc.] Once again, the thematic content of these extracted narratives are consistent with the predefined list of seed words.

To further evaluate whether the sLDA-based topics are aligned with the narratives discussed by Shiller (2019), we conduct the following analysis. First, for each narrative, we simply count the occurrences of the seed words in the NYT articles. This approach is equivalent to the dictionary-based sentiment estimation commonly used in the literature (see Loughran and McDonald (2016)). Then, similar to the sLDA approach, we weight each topic count by the length of each article and compute the monthly weight for each topic. Finally, for each topic, we compute the pairwise correlation between the sLDA-based weight and the manually counted weight and report the results in Table 2, Panel A. Accordingly, for Panic, the correlation between sLDA-based and simply-counted weights is 69% (significant at the 1% level). Among other topics, pairwise correlations for Stock and Real Estate are 32% and 26% (both significant at the 1% level).

4.2 Summary Statistics

We report the summary statistics for the 10 topic weights in Table 2. As mentioned above, Panel A shows the pairwise correlations between the sLDA-based and frequency-based topic weights, of which the most reliable one is Panic. Panel B reports the first and second moments. For each news article, we estimate a vector of topic weights for the 10 topics via topic modeling (i.e., θ_d^i in the sLDA model in Section 2.3). We then compute the monthly topic weight by averaging the article weights during that month, weighted by the length of the article. Accordingly, Panic on average receives the most attention with a mean time-series weight of 11.92%. Intuitively, 11.92% of the monthly NYT articles use at least one of the Panic words at least once.

Table 2 shows Panic is also the second-most volatile topic with a standard deviation of 4.13% after Saving at 4.41%. Real estate attracts the least attention with a mean of 7.54% and is also the least volatile with a standard deviation of 2.62%. Panel C shows the autocorrelations in which Stock is the most persistent narrative with a first-order autocorrelation of 82.12%. Panic comes in second at 78.19%.

In the following empirical analyses on stock market implications of news, in addition to considering stand-alone narratives, we also create a composite narrative index by extracting and combining the signals most relevant to return prediction from all topics via the two-step PLS method, which has recently gained wide popularity in the literature (Kelly and Pruitt, 2013, 2015; Huang et al., 2015, 2020). Specifically, as a first step, the time series of each topic weight is regressed on the time series of next-period market returns using the whole sample. Second, in each period t, the vector of topic weights is regressed on the vector of slopes obtained in the first step. The slope in the second step regression is a value of the PLS index in period t. We note that the construction of the PLS index for in-sample analyses uses the full sample from 1871 to 2019 in the same spirit as Huang et al. (2015) and Huang et al. (2020). For the out-of-sample analysis, to avoid any look-ahead bias, we recursively reconstruct the PLS index every month using only the information available up to that month.

Panel D of Table 2 reports the PLS loadings (the slope in the time-series regressions) for all topics.[14] As expected, Panic receives the highest weight, and its positive loading indicates that Panic is a positive predictor of market returns. The second-most important topic is Boycott, which has a negative PLS weight. Among the other topics, Confidence, Consumption, and Wage have positive weights, whereas the remainder display negative weights.

Panel E of Table 2 reports the correlations among the 10 topics and the PLS index. As expected, the PLS index is highly correlated with Panic with a correlation coefficient of 81%; its correlation coefficient with Boycott is -53%, while its correlations with the remaining topics are at most 25% in absolute terms. These correlations indicate that the PLS index, while a stronger market predictor, retains all the forecasting features of Panic.

4.3 Time Series of Economic Narratives

Next, we examine fluctuations in topic weights over time. Specifically, we plot the time series of each topic weight against excess market returns from January 1871 to October 2019. Both have been demeaned for ease of visualization. The results for the five main

[14]By nature of construction, the absolute PLS weights do not carry much meaning. We only care about the relative weights of the components.

topics are displayed in Figure 4, and the remaining topics are displayed in Figure B2 in the Internet Appendix.

First, as Panic and the narrative PLS are highly correlated, their plots look nearly identical. As shown in the word cloud in Figure 3, Panic encompasses a number of stress-related themes, including economic crises, epidemics, and particularly, international tensions and political risks. Tracking the time series of Panic in Figure 4, one can see that Panic spiked in the 1870s, a period of settlements and reconstructions after the American Civil War. It also surged during the 1890s, an eventful period featuring the Panic of 1893, the Spanish-American War in 1898, and the Philippine-American War of 1899–1902. The Panic index rose to its highest since the start of the sample during the World War I from 1917 to 1918. It remained low during the 1920s and 1930s before surging again during the World War II. Panic reached its all-time high in 1963, the year of the assassination of President John F. Kennedy.

As will be shown later in this paper, the impact of Panic on the stock market is increasing over time. Hence, in Figure 5, we zoom in on the time series of Panic over the past 30 years. We track down the 20 articles with the biggest contribution to the 20 highest monthly scores of Panic since 1990.[15] Over the past 30 years, Panic spiked in the early 1990s, during the Gulf War, and surged again at the end of 2001, after the 9/11 terrorist attack. During recent years, Panic remains high, especially from 2014 to 2018. The most important articles during this time reflect the period's climate: stories are full of international tension, most notably the war in the Middle East, the nuclear weapon threat from North Korea, and the alleged meddling of Russia in the 2016 Presidential Election. Overall, stories about international tensions and political risks have clearly contributed the most to Panic over the past 30 years.

Another narrative of particular interest is Stock Market Bubbles. Over the span of 150 years, Stock, unlike Panic, does not display a clear pattern of ups and downs. As shown in Figure 4, Stock was high during World War I and the 1920–1921 deflationary recession. It spiked up in 1929—at the onset of the Great Depression—and during the recession from 1937 to 1938. It was on an upward trend from the mid-1940s to the mid-1960s and fluctuated wildly during the 1970s and 1980s. Figure 5 zooms in on the times series of Stock over the past 30 years along with the most influential articles contributing to each spike. In the mid-1990s, a lot of attention was paid to the unprecedented rise in uncovered

[15]The most influential article each month is the article with the highest product of article-level topic weight and article length. Equal weighing, that is, ignoring an article's length, can help one identify slightly different influential articles, but these slightly different articles are generally thematically similar to the most influential articles reported here.

short sales. The Stock narrative weight surged in early 2000, after the dot-com bubble, and rose to its 50-year height during the 2007–2009 financial crisis. The most important articles during this period mainly address the challenges faced by leading banks, such as Goldman, Bank of America, and Citigroup. The stock market bubble narrative has been on a downward trend during the most recent decade, except for one spike during the stock market sell-off in 2015.

5 Economic Narratives as Stock Market Predictors

In this section, we address the primary research question of the paper: do economic narratives predict the U.S. stock market return? We first consider one-month return predictions, before moving to long-horizon predictions. In the last part of this section, we will control for common return predictors, and we find that narratives have additional predictive power.

5.1 Predicting Next One-Month Returns

To investigate the return predictability of economic narratives, we run the following standard predictive regression:

$$R^e_{t+1} = \alpha + \beta x_t + \epsilon_{t+1}, \tag{2}$$

where R^e_{t+1} is the annualized excess market return over the next month, x_t is one of the narratives or the narrative PLS index standardized to zero mean and unit variance, and β, the coefficient of interest, measures the strength of predictability. The reported t-statistics are computed with Newey and West (1987) standard errors.

Table 3 reports the regression results. Over the whole sample of 1871–2019, among the 10 narratives, Panic is a strong positive predictor with the coefficient being statistically significant at the 1% level. Economically, a one-standard-deviation increase in Panic is associated with a 3.4% increase in the annualized excess return in the next month. In addition to the full sample analysis, we also run predictive regressions over two subperiods and for the last two decades. This approach serves two purposes: First, we want to address concerns about text quality in the earlier sample. Second, the last two decades present a unique period. Internet usage and the use of technology in diffusing information significantly changed after the year 2000, and how they will affect financial markets and society in the future is still unclear. In addition, the results during this period are probably the most relevant for the future as emphasized in Goyal and Welch (2008). Accordingly, the positive association between Panic and future market returns continues to hold in both subperiods with significance at the 10% and 1% levels, respectively. Notably, over the past two decades, Panic yields an impressive forecasting power with a coefficient of 10.9%, significant at the 1% level, and an in-sample R^2 of 4.2%.

Among the remaining economic narratives, only Boycott and Tech produce statistically significant predictive coefficients, at the 5% and 1% levels, respectively, but the results seem to be driven by the subperiod of 1950–1999.

The last portion of Table 3 reports the prediction results with the narrative PLS index. As expected, the PLS index displays the prediction results which are similar to and stronger than Panic. Over the full sample, a one-standard-deviation increase in the PLS index foreshadows a 5.6% increase in the annualized return in the next month. The PLS index displays increasingly stronger predictive results over the subsamples, especially during the latter subperiod and the last two decades, where the results are significant at the 1% level.

Following Golez and Koudijs (2018), we compute the cumulative in-sample R^2 in predicting the next month's returns and report the results in Figure 6. An upward trend indicates a predictor performs well during a period. Accordingly, both Panic and the PLS index experience poor performances during the period 1910–1930 but show a strong recovery after that. Again, both suffer from a slight drawdown for a short period before 2000.

Overall, Table 3 indicates that Panic and the PLS index are strong market predictors, and their forecasting power is increasing over time. The predictability of Panic and the PLS index is most pronounced during 2000–2019. We conjecture that the digitalization of news and the technology that accelerates the diffusion of information drive this result. This result is consistent with that of Obaid and Pukthuanthong (2021), who also find strong market predictability in the sentiment of photos and text after the 2000s.

5.2 Predicting Long-Horizon Returns

In the previous subsection, we find that Panic and the PLS index can predict market returns over the next one month. Hence, in this subsection, we further examine the long-horizon predictability of Panic and the PLS index by running the following predictive regression:

$$R^e_{t+1\rightarrow t+h} = \alpha + \beta x_t + \epsilon_{t+1\rightarrow t+h}, \tag{3}$$

where $R^e_{t+1\to t+h}$ is the annualized excess market return over the next h periods, x_t is either Panic or the PLS index, and β, the coefficient of interest, measures the strength of predictability. To account for potential autocorrelations of the residuals in the long-horizon predictive regressions, we compute Newey and West (1987) standard errors with corresponding h lags. With our large sample (of more than 1,700 monthly observations), and because the first-order autocorrelations of Panic and the PLS index are under 80%, regression (3) is less likely to be subject to the Stambaugh (1999) bias.

The first row of each panel in Table 4 repeats the results for $h = 1$ for comparison. Panel A of Table 4 reports the results for the full sample from 1871 to 2019. Over this 150-year period, both Panic and the PLS index can significantly positively predict market returns up to 36 months ahead.

In the subsample analysis, Panic's predictability is weak during the first half of the sample period but becomes strongly significant at least at the 5% level or better from one to six months during the second subperiod (1950 to 2019). The predictability of the PLS index is significant at the 5% level for the next 1 and 36 months and at the 10% level for other forecasting horizons during the first half of our sample period. The results revert and become strongly significant during the second half.

The most exciting results are obtained from the year 2000. Over the past 20 years, Panic yields impressive predictability: its in-sample adjusted R^2 ranges from 4% (1 month) to 20% (36 months). During this period, the PLS index, unlike Panic, produces stronger one-month results but weaker long-horizon results. However, this pattern is expected as the PLS index is constructed so as to maximize its one-month predictability.

5.3 Predicting Market Returns: Controlling for Economic Variables

The reported predictability of narratives could simply reflect other economic variables as covered by the news articles. To investigate this possibility, we rerun the predictive regression with the addition of common economic predictors as control variables:

$$R^e_{t+1} = \alpha + \beta x_t + \gamma z_t + \epsilon_{t+1}, \tag{4}$$

where z_t is one of the economic predictors. Following Huang et al. (2020), we include as economic predictors the 14 variables from Goyal and Welch (2008), the output gap from Cooper and Priestley (2009), and the short interest index from Rapach et al. (2016).

Panel A of Table 5 reports the results of single predictive regressions when each of the

16 economic variables is used alone to predict the next month's excess market return. As shown in Goyal and Welch (2008), most of these variables are not significant as a market predictor. The Treasury-bill rate and short interest are negative predictors, significant at the 5% level, while the long-term bond return is the only significantly positive predictor, marginally significant at the 10% level. The last row reports the prediction results with a PLS index constructed with all 16 economic variables. Accordingly, the economic PLS index is significant at the 10% level.

In Panel B of Table 5, the narrative PLS index is tested against each economic predictor. The PLS index is used instead of Panic as the former inherits the features of the latter and is a stronger predictor. The PLS index remains significant at the 1% level in the face of the 15 economic predictors. Finally, when tested against the economic PLS index, the narrative PLS index remains significant at the 5% level and drives out the significance of the economic index. Overall, the results in this section highlight that the narratives contain valuable information beyond what is encapsulated by the common economic variables.

5.4 Predicting Market Returns: Controlling for Uncertainty and Sentiment Variables

In the previous section, we documented that the narrative PLS index contains valuable insights into market returns. In this section, we ask whether the narrative index reflects information contained in other well-known uncertainty or sentiment variables. Recently, ample uncertainty measures have been introduced into the literature, notably the financial and macro uncertainty indexes from Jurado et al. (2015), the economic policy uncertainty index from Baker et al. (2016), and the disagreement index from Huang et al. (2020). Another commonly used measure of uncertainty is the Chicago Board Options Exchange's Volatility Index (VIX). Manela and Moreira (2017) develop a measure of VIX as captured by the WSJ frontpages and extend this index to period preceding the existence of VIX.

Another rising strand of the predictability literature studies sentiment measures. Most influential is the investor sentiment index, developed by Baker and Wurgler (2006) (hereafter BW sentiment), and has been documented to have predictability over small and hard-to-value stocks. Huang et al. (2015) extract the components most relevant to market returns from the BW sentiment using the PLS method to construct a powerful predictor (hereafter PLS sentiment). Most recently, Jiang et al. (2019) constructed manager sentiment from corporate filings to show that manger sentiment has predictability beyond what is captured by investor sentiment. Moreover, Tetlock (2007) and Garcia (2013) show that sentiment extracted from news articles can predict daily market returns. To construct news sentiment from the NYT, we simply compute the difference between the percentages of positive and negative words belonging to the sentiment dictionary developed in Loughran and McDonald (2011) (the most well-known sentiment dictionary in finance research). Finally, we use as control variables the two U.S. stock market confidence indexes introduced by Shiller: the one-year confidence index and the crash confidence index.[16]

Panel A of Table 6 reports the pairwise correlation between the narrative index and each uncertainty and sentiment index. The narrative PLS index has a significant 21% correlation with the economic policy uncertainty in Baker et al. (2016) and a significant -18% correlation with the manager sentiment index in Jiang et al. (2019). Furthermore, the narrative PLS index is significantly negatively correlated with Shiller's one-year confidence index at -29%.

Panel B of Table 6 reports the univariate prediction for each uncertainty and sentiment variable. As expected, disagreement, PLS investor sentiment, and manager sentiment are strong negative market predictors as recently documented by the literature. The financial uncertainty index by Jurado et al. (2015) is a negative predictor significant at the 5% level. Notably, the well-known economic policy uncertainty index developed by Baker et al. (2016) is not significant in a one-month regression, consistent with previous studies (see, e.g., Pástor and Veronesi (2013) and Brogaard and Detzel (2015)).

Panel C tests the narrative PLS index against the other sentiment and uncertainty variables. The PLS index remains significant (at least at the 5% level) in each multivariate predictive regression. The last row of Table 6 reports the results with the PLS index constructed from all the uncertainty and sentiment variables, against which the PLS index remains significant at the 1% level.

The results in Table 6 indicate that the narrative index contains valuation information about market returns after controlling for the strong market predictors recently proposed in the literature.

[16]These indexes are available at https://som.yale.edu/faculty-research-centers/centers-initiatives/international-center-for-finance/data/stock-market-confidence-indices/united-states-stock-market-confidence-indices.

6 Panic as a Proxy for Time-Varying Risk Aversion

In Section 5, we show that Panic is a strong market return predictor and its predictive power is increasing over time. Furthermore, the narrative PLS index, which loads strongly on Panic, contains predictive power beyond common economic and uncertainty return predictors. Thus, investigating what specifically Panic captures is an interesting task.

In this section, we first hypothesize that Panic proxies for time-varying risk aversion and then provide empirical evidence in favor of this hypothesis.

6.1 Panic and the Risk-Return Trade-Off

Before hypothesizing that Panic captures time-varying risk aversion, we first briefly introduce the Merton (1973) ICAPM model. In his seminal paper, Merton (1973) derives the following classic risk-return trade-off between the conditional mean of the return on the wealth portfolio, $\mathbb{E}_t[R_{M,t+1} - R_{f,t+1}]$, its conditional volatility, $\sigma^2_{M,t}$, and its conditional covariance with the investment opportunity set, $\sigma_{MF,t}$:

$$\mathbb{E}_t[R_{M,t+1} - R_{f,t+1}] = \left[\frac{-J_{WW}W}{J_W}\right]\sigma^2_{M,t} + \left[\frac{-J_{WF}}{J_W}\right]\sigma_{MF,t}, \tag{5}$$

where $J(W(t), F(t))$ is the indirect utility function in wealth, $W(t)$, and any state variables, $F(t)$, describing the evolution of the investment opportunity set over time. The term $\lambda \equiv \left[\frac{-J_{WW}W}{J_W}\right]$ (subscripts denote partial derivatives) is linked to the measurement of relative risk aversion (RRA) and is expected to be positive. Hence, the first term in equation (5) captures the positive risk-return trade-off in which market participants require a higher risk premium on the wealth portfolio when its payoff is expected to be more uncertain. The second term in equation (5) links the risk premium on the wealth portfolio to innovations in the investment opportunity set. Accordingly, investors will demand a higher risk premium on a wealth portfolio that pays off precisely in states of the world in which the marginal utility of wealth is low. The converse is true when the wealth portfolio serves as a hedge against investment risks.

Following Lundblad (2007) and the majority of papers in this literature, we consider a univariate version of equation (5):

$$\mathbb{E}_t[R_{M,t+1} - R_{f,t+1}] = \lambda_0 + \lambda_1 \times \sigma^2_{M,t}, \tag{6}$$

where we assume that the investment opportunity set is constant or that the representative investor has a log utility function. A natural step then is to empirically test the univariate risk-return trade-off as depicted in equation (6) with the popular GARCH-in-mean framework developed Bollerslev (1986) and Engle and Bollerslev (1986). Specifically, we consider first the following mean equation for the mean-volatility trade-off:

$$R_{M,t+1} - R_{f,t+1} = \lambda_0 + \lambda_1 \times \sigma^2_{M,t} + \epsilon_{t+1}, \tag{7}$$

where ϵ_{t+1} has a mean of zero with conditional variance $\sigma^2_{M,t}$. Empirical tests of equation (7) on the U.S. stock market return has yielded mixed results, depending on the sample period and the specification of the volatility equation. Lundblad (2007) reconciles the contradictory findings on the U.S. risk-return trade-off present in the literature. He employs a long sample of U.S. stock market returns and documents a strong positive trade-off. He notes that a weak empirical relation may be an artifact of small samples and hence emphasizes the use of large samples in studying the risk-return relationship.

The specification in equation (6) and equation (7) assumes that the coefficient of relative risk aversion, λ_1, is time invariant. However, we have no compelling reason to believe that this assumption would hold in practice. Indeed, in asset pricing models, such as the external habit model by Campbell and Cochrane (1999), relative risk aversion is modeled as time varying. If we assume time-varying relative risk aversion, then we can specify the risk-return trade-off as a linear function of some state variable, x_t:

$$R_{M,t+1} - R_{f,t+1} = \lambda_0 + (\lambda_1 + \lambda_2 \times x_t) \times \sigma^2_{M,t} + \epsilon_{t+1}. \tag{8}$$

We hypothesize that Panic proxies for time-varying relative risk aversion, and, thus, we replace the state variable, x_t, with Panic in equation (8). Hence, $\lambda_t = \lambda_1 + \lambda_2 \times Panic_t$. If this hypothesis holds with real-world data, then we expect (1) the adjusted R^2 of equation (8) to be higher than that of (7), as the former is a more proper representation of the risk-return trade-off, and (2) the coefficient λ_2 in equation (8) to be significantly positive as risk aversion is expected to rise when Panic is high.

To complete the GARCH-M framework, we need a specification for the conditional volatility equation. Following Lundblad (2007), we consider four different volatility specifications, namely, GARCH (Bollerslev, 1986), IGARCH (Engle and Bollerslev, 1986), TGARCH (Zakoian, 1994), and EGARCH (Nelson, 1991):

$$\begin{aligned}
\text{GARCH}(1,1): \ & \sigma^2_{M,t} = \delta_0 + \delta_1 \epsilon^2_t + \delta_2 \sigma^2_{M,t-1} \\
\text{IGARCH}(1,1): \ & \sigma^2_{M,t} = \delta_0 + \delta_1 \epsilon^2_t + (1-\delta_1) \sigma^2_{M,t-1} \\
\text{TGARCH}(1,1): \ & \sigma^2_{M,t} = \delta_0 + \delta_1 \epsilon^2_t + \delta_3 D_t \epsilon^2_t + \delta_2 \sigma^2_{M,t-1} \\
\text{EGARCH}(1,1): \ & ln(\sigma^2_{M,t}) = \delta_0 + \delta_1 \left(\frac{|\epsilon_t|}{\sigma_{M,t}}\right) - \delta_3 \left(\frac{\epsilon_t}{\sigma_{M,t}}\right) + \delta_2 ln(\sigma^2_{M,t-1}),
\end{aligned} \tag{9}$$

where D_t is an indicator equal to one when ϵ_t is negative and zero otherwise.

Panel A of Table 7 reports the results using the standard GARCH(1,1) model. Over the whole 150-year sample, the coefficient of RRA, λ_1, is 2.17, significant at the 5% level. Hence, with a large sample size, we observe the positive risk-return trade-off. However, the adjusted R^2 is negative at -0.38% as the conditional volatility is very smooth, failing to explain the variations in realized returns. These results are consistent with those of Lundblad (2007). Moving on to the time-varying RRA specification, if Panic proxies for time-varying RRA, we expect the interaction term λ_2 to be significantly positive and the conditional volatility to have higher explanatory power for return variations. The empirical results in Panel A confirm these conjectures. Specifically, λ_2 is 1.58, significant at the 1% level, and the adjusted R^2 jumps from -0.38% to 0.17%, indicating a better fit. Notably, the coefficient capturing constant RRA, λ_1, collapses toward zero.

We obtain similar results when decomposing the whole 150-year sample into two subsamples as in the previous tests of return predictability. In the first half of the sample, the time-varying RRA specification yields a better model fit as measured by R^2, but the coefficient λ_2 is not significant. In the second subsample, R^2 jumps more than 10 times and λ_2 is highly significant under the time-varying RRA model.

Panels B, C, and D of Table 7 report the results with different specifications for the volatility equation. That we obtain consistent results across both the models and sample periods, except for TGARCH in the first subsample, confirms that Panic captures risk aversion, thereby enhancing the risk-return relationship.

6.2 Panic and Market Volatility

In the previous subsection, we show that Panic captures relative risk aversion and thus improves the risk-return trade-off via estimation of the popular GARCH-M model. Another angle from which to study the relationship between Panic and the risk-return trade-off is to re-examine equation (6), which, ignoring the constant term, can be restated as

$$\frac{\mathbb{E}_t[R_{M,t+1} - R_{f,t+1}]}{\sigma^2_{M,t}} \approx \lambda_t, \tag{10}$$

where λ_t is a measure of time-varying RRA. Equation (10) states that when investors are more risk averse, they demand a higher market risk premium and/or a lower market volatility to participate in the stock market; that is, they require a higher Sharpe ratio to hold the market portfolio. If investors' risk aversion rises when Panic is high, Panic is then a positive predictor for market returns and a negative predictor for market volatility. Section 5 already shows that Panic is a strong market predictor, so in this subsection, we will investigate its predictability over market volatility.

To examine this hypothesis, we conduct the following predictive regression:

$$\sigma_{t+1\to t+h} = \alpha + \beta Panic_t + \delta' W_t + \epsilon_{t+1\to t+h}, \tag{11}$$

where $\sigma_{t+1\to t+h}$ is either the realized or implied market volatility in annualized percentages over the next h months, and W_t is a set of controls. Realized volatility is the square root of the sum of squared daily market returns over the h periods, rescaled to annual values. Following Baker et al. (2016), we simply compute the average of the daily VIX over h periods. Calomiris and Mamaysky (2019) and Glasserman and Mamaysky (2019) document market volatility to be clustered and predictable. Following their studies, we include a set of well-known volatility predictors as controls. When σ_t is realized volatility, W_t includes two lags of realized volatility and two lags of negative market returns. The VIX is only available from 1990, so when VIX is the independent variable, W_t includes two lags of VIX, two lags of realized volatility, and two lags of negative market returns.

Panel A of Table 8 reports the results for the realized market volatility. As we need daily excess returns to compute the realized volatility, the full sample begins from 1927 and ends at 2019. We also examine three subperiods: 1927–1949, 1950–2019, and 2000–2019. As expected, Panic is a statistically strongly negative predictor of the realized market volatility, although the economic significance is modest. For example, over the full sample, a one-standard-deviation increase in Panic predicts a 0.6% and 0.7% decrease in annualized realized volatility over the next one and three months, respectively. Since the turn of the twenty-first century, the economic magnitude becomes stronger. For example, a one-standard-deviation increase in Panic is associated with a 1% decrease in annualized realized volatility over the next three months. Panel B reports results for the VIX over the period 1990–2019. Again, we find that Panic is a strong negative predictor of implied volatility over the next one and three months.

Overall, the empirical results in this and the previous subsections support the hypothesis that Panic proxies for time-varying RRA.[17] A naturally arising question is whether Panic as measured from news articles proxies for time-varying changes in RRA by itself or whether Panic reflects RRA fluctuations caused by real-world events. We will attempt to answer this question in the next subsection by comparing the predictability of Panic with real-time events.

[17]Using the narrative PLS index in place of Panic in the above tests with ICAPM and volatility prediction yields similar results, which, in an effort to conserve space, have not been reported.

6.3 Panic versus Actual Events

We construct Panic to be the attention paid in news articles to various stress-inducing themes, including recessions, bank failures, wars, disasters, and epidemics. To investigate whether Panic or the actual events have predictability over market returns, we first create indicators for these events reported by GFD:[18]

- Recessions: from NBER;
- Bank failures: if the event is tagged as bank failure, panic, or crime;
- Wars: if the event is tagged as war, military, revolution, assassination, rebellion, insurrection, riot, terrorism, battle, or invasion;
- Disasters: if the event is tagged as disaster, earthquake, weather, tornado, hurricane, or typhoon;
- Epidemics: if the event is tagged as epidemic or pandemic;
- All: if the event is tagged with any of the above.

We then include these event indicators as controls in the predictive regressions:

$$R^e_{t+1} = \alpha + \beta \times Panic_t + \gamma^j \times D^j_t + \epsilon_{t+1}, \tag{12}$$

where D^j_t is a dummy variable for event j equal to one if there is one event j in month t. If Panic contains additional predictive power, β is expected to be significantly positive.

Panel A of Table 9 reports the results for the whole sample. Across all events, Panic remains its significance as a return predictor. Among the events, only Recessions yields a significant prediction coefficient, albeit a negative one; hence, the result is inconsistent with a risk-based explanation. The results indicate that the actual events themselves, except Recessions, have no predictive power and thus cannot be a cause of fluctuations in RRA. This evidence rules out the possibility that Panic only reflects RRA changes triggered by real-word stressful events. Panel B reports the results in the first half of the sample from 1871 to 1949. During this period, Panic is marginally significant, except when tested against Recessions and Disasters. During the second half of the sample, Panic remains significant at the 1% level across all events and drives out the significance of Recessions.

Overall, the findings in this subsection eliminate the alternative explanation that the predictability of Panic from news articles is simply a manifestation of actual events. Indeed, we find that most of the events have no predictive power, and those displaying predictability are negative predictors. Thus, the events themselves do not lead to changes

[18] Figure B3 in the Internet Appendix plots these events.

in relative risk aversion, consistent with the theoretical discussion so far.

Alternatively, the market predictability of Panic is also consistent with the prediction of the asset pricing model with rare disaster risks. Specifically, according to Barro (2009), an increased probability of a rare disaster explains increases in the equity premium. As Panic includes various disaster themes, such as wars, political tensions, and epidemics, the implied disaster probability is expected to spike during times when Panic is high. As a result, higher Panic predicts higher future equity premium, confirmed by the empirical findings.

7 Out-of-Sample Analysis

The predictability results in Section 5 are obtained with the whole 150-year sample. To offer real-time investors economic value, return predictors need to have out-of-sample forecasting power (Goyal and Welch, 2008). To investigate whether economic narratives can help investors make better investment decisions, we conduct two standard out-of-sample tests: out-of-sample R^2 and certain equivalent return (CER) gains.

7.1 Out-of-Sample R^2

Following Campbell and Thompson (2008), we compute the following well-known out-of-sample R^2 statistic:

$$R^2_{OS} = 1 - \frac{\sum_{t=p}^{T-1} \left(R^e_{t+1} - \hat{R}^e_{t+1} \right)^2}{\sum_{t=p}^{T-1} \left(R^e_{t+1} - \bar{R}^e_{t+1} \right)^2}, \tag{13}$$

where R^e_{t+1} is the realized excess market return, $\hat{R}^e_{t+1} = \hat{f}_t(x_t)$ is the predicted excess return with $\hat{f}_t(x_t)$ being a function of the predictors recursively estimated using only the training window, $\bar{R}^e_{t+1}$ is the historical mean excess return computed over the training window, and p is the size of the initial training window. We employ an expanding estimation window to incorporate all available information into formulating future forecasts and begin the evaluation period in January 1891 (20 years from the start of the sample) to guard against any concerns about window size manipulation.

We benchmark the out-of-sample results of the 10 narratives against the six economic predictors from Goyal and Welch (2008), including dividend-price ratio, dividend yield, earnings-price ratio, dividend payout ratio, stock variance, and Treasury-bill rate, all of which are available from 1871.

We use three approaches to recursively estimate the function $f_t(x_t)$. First, $f_t(x_t)$ is a linear function of each of the 10 topics and each of the 6 economic predictors. Second, $f_t(x_t)$ is a function of either all 10 narratives or all 6 economic predictors estimated via PLS as described in Section 4.

Finally, $f_t(x_t)$ as a function of either all 10 narratives or all 6 economic predictors is estimated via random forest, an advanced machine learning method developed by Breiman (2001). Intuitively, random forest is an ensemble method averaging predictions from de-correlated trees where each tree is a partition of the feature space (predictors) onto a set of nodes (leaves) and then fits a constant to each node. Random forest is ideal in this context because it requires minimal tuning of hyperparameters and generally is not subject to overfitting, a common phenomenon in other advanced machine learning methods (Hastie et al., 2009, chap. 15).[19] Also, recall that our topic weights are extracted every month using data over the past 10 years, so there is no look-ahead bias in the out-of-sample analysis.

If a predictor outperforms the historical mean benchmark in forecasting future returns, it will produce a smaller mean squared forecast error (MSFE) than that of the historical mean, and, thus, the R^2_{OS} will be greater than zero. To test the significance of R^2_{OS}, we report the Clark and West (2007) MSFE-adjusted statistic.

Panel A of Table 10 reports the results from an OLS regression using individual predictors. Among the six economic predictors, only Treasury Bill produces a positive and significant R^2_{OS} over the whole evaluation period, yet the magnitude is tiny at 0.07%. Meanwhile, among the 10 narratives, during 1891–2019, Panic and Boycott are the only predictors to yield a significant R^2 (0.28% and 0.15%, respectively). These three predictors also deliver out-of-sample predictions over the period of 1950–2019, while Panic is the only predictor among all those considered to produce a positive R^2 in the most recent sample, at 1.41%, significant at the 1% level. Consistent with the in-sample results in Section 5, Panic displays strong out-of-sample predictability, and its predictive power is increasing over time.

Panel B of Table 10 combines the signals of individual predictors via PLS. The combination of all six economic predictors produces negative R^2's across all sample periods. In contrast, combining all narratives via PLS yields an R^2 of 0.24%, significant at the 5% level. However, this value is smaller than that produced by stand-alone Panic in Panel A. In the two most recent subsamples, the narrative PLS method improves on Panic,

[19] We use the package randomForest in R and the out-of-bag (OOB) error to select the optimal number of features during each estimation period.

producing R^2's of 1.07% over 1950–2019 and 1.71% over 2000–2019, significant at the 1% and 10% levels, respectively. In Panel C, using random forest on either economic predictors or narratives results in insignificant R^2's across all sample periods. However, as will be shown in the next subsection, the use of random forest leads to better asset allocation results.

We also consider a rolling window estimation scheme with 600 months and report the results in Table B2 in Internet Appendix B. Compared with the expanding window used in Table 10, Panic yields higher R^2_{OS} during both the whole sample and the past 20 years. From 2000 to 2019, Panic produced an R^2_{OS} of 2.6%, significant at the 1% level. The combination of all topics via PLS also yields better results over the whole window but worse results during the past two decades.

Figure 7 plots the cumulative out-of-sample R^2 for Panic and the PLS method for all narratives. The results for random forest are not displayed, because random forest yields a large negative R^2_{OS} before 2000, making it difficult to visualize the trends of the other two methods. An upward trend indicates favorable performance during that period. Consistent with Table 10, Panic displays a steadily upward trend over the evaluation sample with a slight drawdown before 2000. Meanwhile, PLS displays a much more volatile pattern for R^2_{OS} that rapidly rises during the first 20 years of the sample, before collapsing during the next 30 years. The PLS R^2_{OS} turns around during the second half of the twentieth century, peaks in the 2007–2009 financial crisis, and deteriorates over the most recent decade.

Overall, the analysis in this subsection highlights the outperformance of Panic over common economic predictors in out-of-sample prediction exercises, confirming that Panic is a robust return predictor, both in sample and out of sample. Plus, using other narratives improves the out-of-sample predictability of Panic during recent periods.

7.2 Asset Allocation Implications

In this section, we further examine the economic value of news narratives from an asset allocation perspective. Following Campbell and Thompson (2008), we compute the certainty equivalent return (CER) gain and Sharpe ratio for an mean-variance investor who optimally allocates her portfolio between the stock market and a risk-free security using out-of-sample return forecasts.

At the end of period t, the investor optimally allocates

$$w_{t+1} = \frac{1}{\gamma} \frac{\hat{R}^e_{t+1}}{\hat{\sigma}^2_{t+1}}, \tag{14}$$

of the portfolio to equities during period $t+1$, where γ is the risk aversion coefficient set to three following Huang et al. (2020),[20] $\hat{R}^e_{t+1}$ is the predicted excess return, and $\hat{\sigma}^2_{t+1}$ is the variance forecast. The investor then allocates $1 - w_{t+1}$ of the portfolio to the risk-free asset. The $t+1$ realized portfolio return is

$$R^p_{t+1} = w_{t+1} R^e_{t+1} + R^f_{t+1}, \tag{15}$$

where R^f_{t+1} is the risk-free return. Following Campbell and Thompson (2008), we use a rolling window of 60 months to estimate the variance forecast of the excess market returns and constrain w_t between 0 and 1.5 to exclude short sales and allow a maximum 50% leverage.

The CER of the portfolio is

$$CER_p = \hat{\mu}_p - 0.5\gamma\hat{\sigma}^2_p, \tag{16}$$

where $\hat{\mu}_p$ and $\hat{\sigma}^2_p$ are the sample mean and variance, respectively, for the realized portfolio returns over the evaluation period. The CER gain is then the difference between the CER for an investor who uses a forecasting model to predict the excess market return and the CER for an investor who uses the historical mean forecast. We annualize the CER gain by multiplying by 12 so that it can be interpreted as the maximum management fee the investor is willing to pay to gain access to the predictive forecasts. In addition to the CER gain, we also compute the annualized monthly Sharpe ratios of the portfolio's realized returns. We test the statistical significance of the CER gain and the Sharpe ratios (against the historical mean benchmark) using the test statistics in DeMiguel et al. (2009).

Panel A of Table 11 reports the asset allocation results when individual predictors are used to make return forecasts. Among all predictors, Treasury Bill produces the highest utility gains during the whole sample of 1891–2019 at 1.43%, while Panic comes in second at 0.56%. While Panic makes better return forecasts as indicated by a higher R^2_{OS} (reported in Table 10), its forecasts may be much more volatile than those of Treasury Bill, which may explain the outperformance of Treasury Bill in terms of utility gains. The earning-price ratio produces a low utility gain over the whole sample but yields an

[20] To conserve space, the results with a risk aversion coefficient of five are not reported but are similar to the reported results and are available on request.

impressive value of 3.88% in the most recent period of 2000–2019.

Panel B of Table 11 shows that the combination of narratives via PLS produces a utility gain three times as big as that of the economic predictors over the whole sample, while the two combinations yield similar results in the most recent period. In Panel C, random forest performs better on narratives than on economic predictors and also outperforms PLS on narratives. Particularly, using random forest on narratives over the period of 2000–2019 results in a utility gain of 4.45%, the highest achieved among all forecasts.

The right panel of Table 11 shows the results for the annualized Sharpe ratio. Among all the individual predictors in Panel A, once again Treasury Bill yields the best results for the whole sample, followed by Panic and EP. When we combine all the narratives via either PLS or random forest, we observe an improvement in Sharpe ratio over the use of Panic alone in most of the samples. The last row reports the annualized monthly Sharpe ratio from buying and holding the S&P 500 index in the corresponding periods. Panic or a combination of narratives clearly outperforms the buy-and-hold strategy in general.

This analysis shows that real-time investors, for example, the investor who is willing to pay up to a 4% annual management fee to gain access to the prediction model using random forest over the past two decades, can utilize economic narratives to realize economic gains. Here, the results are consistent with the R^2_{OS} analysis in that the importance of news narratives is increasing over time, especially over the past 20 years.

7.3 Subperiod Predicting Power

In this subsection, we will investigate the predictive power of narratives during different subsamples: expansion versus recession, and high versus low sentiment. The literature seems to have reached a consensus that sentiment indexes can better predict the market during recessionary times (see, e.g., Garcia (2013), Huang et al. (2015), Jiang et al. (2019), among others). The intuition underlying this view is that the fear and anxiety investors feel related to the economic hardships during recessions increase their sensitivity to sentiment (Garcia, 2013).

The literature also shows that sentiment indexes have stronger predictability during high sentiment periods, when mispricings are likely to occur because of short-sale constraints (Stambaugh et al., 2012; Huang et al., 2015; Jiang et al., 2019). Huang et al. (2020) find that their disagreement index yields stronger predictability when sentiment is high: high disagreement leads to higher average bias and more overvaluation. This effect is stronger when investors are more optimistic (Huang et al., 2020). While these observations lean toward the behavioral channel, the predictability of our narratives is more risk based, so whether we can observe similar subsample concentrations in predictability remains unclear.

To examine the above question, we follow Rapach et al. (2010) and Huang et al. (2015), among others. We compute the subsample R^2 as follows:

$$R_c^2 = 1 - \frac{\sum_{t=1}^{T} I_t^c \left(\hat{\epsilon}_t\right)^2}{\sum_{t=1}^{T} I_t^c \left(R_t^e - \bar{R}^e\right)^2}, \quad c = exp,\ \ rec,\ \ high,\ \ low, \tag{17}$$

where I_t^c is an indicator that takes a value of one when month t is an expansion (recession) period or high (low) sentiment period; $\hat{\epsilon}_t$ is the fitted residual based on the in-sample predictive regression (2); $\bar{R}^e$ is the full sample mean of the excess market return; and T is the number of observations for the full sample of 1871–2019. We classify months into expansions and recessions based on National Bureau of Economic Research (NBER) business cycles. For sentiment periods, we follow Stambaugh et al. (2012) and Huang et al. (2015) and classify a month as high (low) sentiment if the Baker and Wurgler (2006) investor sentiment level in the previous month is above (below) is median value for the sample. Unlike the full sample R^2, the subsample R^2 can be positive or negative.

In the same spirit as equation (17), we compute the out-of-sample R_{OS}^2 for each period. Similar to the previous out-of-sample analysis, we use the expanding estimation window, and the evaluation period begins in January 1891.

Panel A of Table 12 reports the results with the in-sample R^2. Accordingly, both Panic and the PLS index yield higher R^2's during recessions (0.67% in recessions vs. 0.11% in expansions for Panic, and 1.03% in recessions vs. 0.90% in expansions for PLS). These results are consistent with observation of concentrated predictive power during recessions documented in the literature. However, when it comes to the out-of-sample R^2 with an expanding window in Panel B, both Panic and the PLS index have stronger predictive power in expansions (0.51% in expansions vs. 0.00% in recessions for Panic, and 0.81% in expansions vs. -0.47% in recessions for PLS). In sum, the results for whether narratives have stronger prediction power in recessions remains inconclusive.

Regarding predicting the performance of narratives across high versus low sentiment periods, we consistently find that narratives can better predict market during low sentiment periods for both in-sample and out-of-sample analyses. For example, the in-sample R^2 for Panic is 0.69% during low sentiment periods versus 0.08% during high sentiment periods, while the figure for PLS is 2.09% versus 0.55%, respectively. For out-of-sample prediction, Panic yields an R^2_{OS} of 0.64% during low sentiment months versus 0.06% during high sentiment months, while the numbers for PLS are 1.18% and 0.04%. While this result is contradictory to the sentiment literature, it is intuitive. When people are in bad mood, they are more receptive to stressful/panic news.

In short, while we do not find evidence of different predicting powers of narratives across the business cycles as commonly documented in the literature, we document that narratives can better predict the market during low sentiment periods. This result is opposite to the sentiment literature. This further indicates that economic narratives predict market outcomes via a different channel from sentiment.

8 Economic Narratives from the *WSJ*

In this section, we check whether narratives predict stock market returns using the topics extracted from about two million WSJ articles over the period 1889–2019. We apply the same estimation method described in Section 2 to obtain the 10 time series of topic weights from the WSJ data. Internet Appendix C reports plots, figures, and summary statistics for the WSJ topic weights.

Table C2 in Internet Appendix C reports the whole sample results in predicting the excess market returns one month ahead using all WSJ narratives. Accordingly, the most influential narrative from the WSJ is Stock Market Bubble. It can negatively predict the market returns over the whole sample. Its predictability is pronounced only the second half of the sample period, although it is weak at 10% significance level. The past two decades paint different pictures for the predictability of Stock. Its predictability is strong and negative at 1% significance level. In contrast to the NYT, Panic from the WSJ, on the other hand, is a weaker positive predictor, only showing significant results (at 5%) over the post-2000 period. The narrative PLS index, which loads negatively on Stock Bubble (biggest weight) and positively on Panic, is hence a positive predictor of market returns.

Consistent with the results from the NYT, the predicting power of the WSJ narratives is also concentrated in the past two decades. We report in Table 13 the long-horizon prediction using the WSJ narratives from January 2000 to October 2019 (the same end month as the NYT period). During this period, the predictability of Panic is stronger over long horizons. In contrast, Stock Bubble extracted from WSJ is a strong negative predictor of market returns over the next one month and its significance remains for up to 12 months ahead.

As both Panic from the NYT and Stock Bubble from the WSJ display strong predictability over the past 20 years, it is interesting to investigate their OOS results using data over the 20-year subsample only. Table 14 reports the R^2_{OS} of Stock from the WSJ compared with that of Panic from the NYT. To guard against data mining in this small sample (20 years), we compute the average R^2_{OS} across different choices of initial training windows, ranging from 60 to 180 months with an increment of 12 months. As reported in Table 14, the average R^2_{OS} for the next one-month return prediction produced by WSJ Stock is 2.72%, more than double the value of Panic at 1.12% (both are significant at at least the 5% level). However, at longer OOS horizons, NYT Panic is more powerful.

Interestingly, the most influential narrative from NYT, Panic, presents risk aversion, while Stock presents mispricing and shows signs of reversal.[21] These findings highlight the importance of the media source in determining narratives. Simply put, the media plays an important role in market activity and appears to customize their narratives to their readers, thereby having a diverse effect on the market.

9 Robustness Tests

9.1 Predicting Returns on Characteristic Portfolios

Previously, we found that Panic and the narrative index can predict market returns and volatility via the channel of risk aversion. These results should hold at the individual portfolio level: risk-averse investors require a higher risk premium to invest in characteristics-based portfolios. Thus, if Panic and the narrative PLS index operate via the risk channel, we expect both to positively predict the excess returns on the portfolios.[22]

Following Huang et al. (2015), we consider 40 characteristics-sorted portfolios, including 10 industry portfolios, 10 size portfolios, 10 book-to-market (BM) portfolios, and 10 momentum portfolios. The sample period for this analysis is from January 1927 to October 2019.

To examine the predictability of narratives over the risk premium on the characteristics portfolios, we run the following predictive regression:

$$R^e_{i,t+1} = \alpha_i + \beta_i x_t + \epsilon_{i,t+1}, \quad i = 1, \ldots, 40 \tag{18}$$

[21] Because of limited space, we do not report the results but make it available on request.
[22] We use the NYT narratives throughout Section 9.

where $R^e_{i,t+1}$ is the excess return on portfolio i, and x_t is either Panic or the PLS index.

Panel A of Table B3 in Internet Appendix B reports results with 10 industry portfolios. Both Panic and the PLS index yield positive slope coefficients across industries, affirming the claim that they proxy for risk aversion. However, different industries have varying degrees of exposure to narratives. For example, Panic can significantly predict returns on only Nondurable, Durable, Technology, and Shopping and Other industries. On the other hand, the PLS index can significantly predict returns on all industries with the strongest predicting powers found in Durable.

The rest of Table B3 reports results with the size, BM, and momentum portfolios. Both Panic and the PLS index yield almost all significant positive slopes for these portfolios. The slopes on the 10 size portfolios increase monotonically from the large to small portfolios for both Panic and the PLS index. The narratives also better predict value (high BM) and past loser stocks. Thus, returns on stocks that are small, distressed (high BM), and recently underperforming are more sensitive to Panic.

9.2 Predicting Market Returns at a Daily Frequency

The empirical analyses up to this point have been performed using the monthly interval. This section will investigate the impact of Panic and narratives on market returns at the daily frequency. To compute daily topic weights, following the monthly approach, we take a weighted average of the article-level weights across all articles during the day, weighted by the number of ngrams in each article. Recall that the topic weights are estimated from the NYT articles at the end of each month, so the daily predictive analysis may be subject to look-ahead bias. Hence, we refrain from conducting any out-of-sample analysis with the daily frequency. However, the daily results are still of interest to practitioners: with sufficient computing power, our estimation scheme can be conducted daily to produce unbiased daily topic weights.

To examine the predictability of Panic and the PLS index constructed with daily topic weights over daily market returns, we run the following predictive regression:

$$R^e_{t+1\to t+h} = \alpha + \beta x_t + \delta' W_t + \epsilon_{t+1\to t+h}, \tag{19}$$

where $R^e_{t+1\to t+h}$ is the excess market return over the next h days, x_t is either Panic or the PLS index constructed from 10 topics, and β, the coefficient of interest, measures the strength of predictability. As is standard in the literature on daily stock market predictions, we also include a vector of controls. Following Tetlock (2007) and Garcia (2013), we use as controls five lags of excess market return, five lags of squared excess return, five lags of sentiment,[23] and weekday indicators.

Panel A of Table B4 in Internet Appendix B reports results over the whole sample period of January 1927 to October 2019 when daily returns are available. We consider return prediction over the next one day and next five days. For one-day returns, a one-standard-deviation increase in Panic predicts an increase of 4.23% in annualized excess returns the next day (significant at the 5% level) and an annualized increase of 3.96% over the next five days (significant at the 1% level). With the PLS index, the corresponding annualized increases are 6.41% and 4.72% (both are significant at the 1% level). This evidence confirms that the predictive power of Panic and the PLS index also holds at the daily frequency.

Panels B and C report the subsample results. Consistent with the monthly results, the predictability of narratives is increasing over time. The most interesting results are obtained for the past 20 years. For example, a one-standard-deviation increase in Panic (the PLS index) predicts an increase of 9% (10%) in annualized excess market returns the next day.

10 Conclusion

In this paper, we build on the economic narratives theory of Shiller (2019). We employ an advanced natural language processing tool called sLDA to extract narratives from nearly two million and seven million *Wall Street Journal* and *New York Times* articles, respectively, over the past 150 years. We create a list of topic-based seed words to input into the sLDA model to guide the topic extraction process. The rolling estimation scheme is designed to include only historical news data at every time of estimation. This approach avoids any look-ahead bias in constructing monthly narrative weights.

Among the narratives considered, the most important one from NYT is Panic, which encompasses various themes related to stress and anxiety. We find that Panic and an index constructed from all narratives are strong market predictors. Specifically, they can positively predict market returns and negatively predict market volatility, both realized and implied. If risk aversion was to rise during panicked and stressful times as captured by the Panic index, the empirical results then would be consistent with a univariate version of the ICAPM of Merton (1973). We find the predictive power of Panic is increasing over time and holds at both market and portfolio levels, as well as at both monthly and daily

[23] As mentioned above, we construct news sentiment using the Loughran and McDonald (2011) sentiment dictionary.

intervals.

That Panic can positively predict excess market returns is also consistent with the literature on rare disaster risks. Specifically, Barro (2009) shows that the probability of rare disasters can explain the high equity premium. As such probability is expected to rise during panicked and stressful times, an increase in Panic is associated with a rise in the equity premium. The empirical findings in this paper confirm this hypothesis.

As a robustness check, we extract narratives from WSJ, and the most important narrative is Stock Market Bubble. Stock Market Bubble is a negative stock market predictor. Stock Market Bubble seems to capture mispricing, as we observe a reversal of stock returns. Our paper sheds light on the importance of considering media sources and their audience when we interpret narratives.

The estimation scheme in this paper could be extended along several dimensions. First, practitioners with enough computing power could estimate our model at a daily frequency to produce unbiased daily topic weights. Second, the estimation scheme also could be adapted to other countries by using local newspapers and translating the seed words into foreign languages.

References

Baker, M. and J. Wurgler (2006). Investor sentiment and the cross-section of stock returns. *Journal of Finance 61*(4), 1645–1680.

Baker, S. R., N. Bloom, and S. J. Davis (2016). Measuring economic policy uncertainty. *The Quarterly Journal of Economics 131*(4), 1593–1636.

Barro, R. J. (2006). Rare disasters and asset markets in the twentieth century. *The Quarterly Journal of Economics 121*(3), 823–866.

Barro, R. J. (2009). Rare disasters, asset prices, and welfare costs. *American Economic Review 99*(1), 243–64.

Berkessel, J., T. Ebert, J. Gebauer, T. Jonsson, and S. Oishi (2021). Pandemics Initially Spread Among People of High (not Low) Social Status: Evidence from COVID-19 and the Spanish Flu.

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

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics 31*(3), 307–327.

Breiman, L. (2001). Random forests. *Machine learning 45*(1), 5–32.

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

Brown, N. C., R. M. Crowley, and W. B. Elliott (2020). What are you saying? Using topic to detect financial misreporting. *Journal of Accounting Research 58*(1), 237–291.

Brown, S., W. Goetzmann, and S. Ross (1995). Survival. *Journal of Finance*.

Bybee, L., B. T. Kelly, A. Manela, and D. Xiu (2021). The structure of economic news. Technical report, National Bureau of Economic Research.

Calomiris, C. W. and H. Mamaysky (2019). How news and its context drive risk and returns around the world. *Journal of Financial Economics 133*(2), 299–336.

Campbell, J. Y. and J. H. Cochrane (1999). By force of habit: A consumption-based explanation of aggregate stock market behavior. *Journal of Political Economy 107*(2), 205–251.

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

Chava, S., A. Hsu, and L. Zeng (2020). Does history repeat itself? Business cycle and industry returns. *Journal of Monetary Economics 116*, 201–218.

Choudhury, P., D. Wang, N. A. Carlson, and T. Khanna (2019). Machine learning approaches to facial and text analysis: Discovering CEO oral communication styles. *Strategic Management Journal 40*(11), 1705–1732.

Clark, T. E. and K. D. West (2007). Approximately normal tests for equal predictive accuracy in nested models. *Journal of econometrics 138*(1), 291–311.

Cooper, I. and R. Priestley (2009). Time-varying risk premiums and the output gap. *The Review of Financial Studies 22*(7), 2801–2833.

DeMiguel, V., L. Garlappi, and R. Uppal (2009). Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy? *The review of Financial studies 22*(5), 1915–1953.

Dimson, E., P. Marsh, and M. Staunton (2009). *Triumph of the Optimists: 101 Years of Global Investment Returns*. Princeton University Press.

Dyer, T., M. Lang, and L. Stice-Lawrence (2017). The evolution of 10-K textual disclosure: Evidence from Latent Dirichlet Allocation. *Journal of Accounting and Economics 64*(2-3), 221–245.

Engle, R. F. and T. Bollerslev (1986). Modelling the persistence of conditional variances. *Econometric reviews 5*(1), 1–50.

Eshima, S., K. Imai, and T. Sasaki (2020). Keyword Assisted Topic Models. *arXiv preprint arXiv:2004.05964*.

Frankel, R. M., J. N. Jennings, and J. A. Lee (2021). Disclosure sentiment: Machine learning vs dictionary methods. *Management Science, forthcoming*.

Gabaix, X. (2012). Variable rare disasters: An exactly solved framework for ten puzzles in macro-finance. *The Quarterly journal of economics 127*(2), 645–700.

Garcia, D. (2013). Sentiment during recessions. *Journal of Finance 68*(3), 1267–1300.

Gentzkow, M. and J. Shapiro (2010). What drives media slant? Evidence from U.S. daily newspapers. *Econometrica*.

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

Golez, B. and P. Koudijs (2018). Four centuries of return predictability. *Journal of Financial Economics 127*(2), 248–263.

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

Griffiths, T. L. and M. Steyvers (2004). Finding scientific topics. *Proceedings of the National academy of Sciences 101*(suppl 1), 5228–5235.

Hassan, T. A., S. Hollander, L. van Lent, and A. Tahoun (2019). Firm-level political risk:

Measurement and effects. *The Quarterly Journal of Economics 134*(4), 2135–2202.

Hastie, T., R. Tibshirani, and J. Friedman (2009). *The elements of statistical learning: data mining, inference, and prediction.* Springer Science & Business Media.

Hillert, A. and M. Ungeheuer (2019). The Value of Visibility. *SSRN*.

Huang, D., F. Jiang, J. Tu, and G. Zhou (2015). Investor sentiment aligned: A powerful predictor of stock returns. *The Review of Financial Studies 28*(3), 791–837.

Huang, D., J. Li, and L. Wang (2020). Are Disagreements Agreeable? Evidence from Information Aggregation. *Journal of Financial Economics*.

Jagarlamudi, J., H. Daumé III, and R. Udupa (2012). Incorporating lexical priors into topic models. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, pp. 204–213. Association for Computational Linguistics.

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

Julliard, C. and A. Ghosh (2012). Can rare events explain the equity premium puzzle? *The Review of Financial Studies 25*(10), 3037–3076.

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

Kelly, B. and S. Pruitt (2013). Market expectations in the cross-section of present values. *The Journal of Finance 68*(5), 1721–1756.

Kelly, B. and S. Pruitt (2015). The three-pass regression filter: A new approach to forecasting using many predictors. *Journal of Econometrics 186*(2), 294–316.

Loughran, T. and B. McDonald (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *Journal of Finance 66*(1), 35–65.

Loughran, T. and B. McDonald (2016). Textual analysis in accounting and finance: A survey. *Journal of Accounting Research 54*(4), 1187–1230.

Loughran, T. and B. McDonald (2020). Textual Analysis in Finance. *Annual Review of Financial Economics 12*(1), 357–375.

Lu, B., M. Ott, C. Cardie, and B. K. Tsou (2011). Multi-aspect sentiment analysis with topic models. In *2011 IEEE 11th international conference on data mining workshops*, pp. 81–88. IEEE.

Lundblad, C. (2007). The risk return tradeoff in the long run: 1836–2003. *Journal of Financial Economics 85*(1), 123–150.

Manela, A. and A. Moreira (2017). News implied volatility and disaster concerns. *Journal of Financial Economics 123*(1), 137–162.

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

Mullainathan, S. and A. Shleifer (2005). The market for news. *American Economic Review*.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica: Journal of the Econometric Society*, 347–370.

Newey, W. K. and K. D. West (1987). A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica 55*(3), 703–708.

Obaid, K. and K. Pukthuanthong (2021). A picture is worth a thousand words: Measuring investor sentiment by combining machine learning and photos from news. *Journal of Financial Economics*.

Pástor, L. and P. Veronesi (2013). Political uncertainty and risk premia. *Journal of financial Economics 110*(3), 520–545.

Rapach, D. E., M. C. Ringgenberg, and G. Zhou (2016). Short interest and aggregate stock returns. *Journal of Financial Economics 121*(1), 46–65.

Rapach, D. E., J. K. Strauss, and G. Zhou (2010). Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. *The Review of Financial Studies 23*(2), 821–862.

Rietz, T. A. (1988). The equity risk premium a solution. *Journal of Monetary Economics 22*(1), 117–131.

Roberts, J. D. and S. O. Tehrani (2020). Environments, behaviors, and inequalities: Reflecting on the impacts of the influenza and coronavirus pandemics in the United States. *International journal of environmental research and public health 17*(12), 4484.

Schwert, G. W. (1990). Stock market volatility. *Financial Analysts Journal 46*(3), 23–34.

Shiller, R. (2005). *Irrational exuberance*. Princeton University Press, Princeton, NJ.

Shiller, R. J. (2017). Narrative economics. *American Economic Review 107*(4), 967–1004.

Shiller, R. J. (2019). *Narrative economics: How stories go viral and drive major economic events*. Princeton University Press.

Stambaugh, R. F. (1999). Predictive regressions. *Journal of Financial Economics 54*(3), 375–421.

Stambaugh, R. F., J. Yu, and Y. Yuan (2012). The short of it: Investor sentiment and anomalies. *Journal of Financial Economics 104*(2), 288–302.

Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *Journal of Finance 62*(3), 1139–1168.

Wachter, J. A. (2013). Can time-varying risk of rare disasters explain aggregate stock market volatility? *The Journal of Finance 68*(3), 987–1035.

Watanabe, K. and Y. Zhou (2020). Theory-Driven Analysis of Large Corpora: Semisupervised Topic Classification of the UN Speeches. *Social Science Computer Review*.

Zakoian, J.-M. (1994). Threshold heteroskedastic models. *Journal of Economic Dynamics and control 18*(5), 931–955.

Figure 1. Estimation Scheme

This figure plots the rolling estimation scheme for the sLDA model. Every month t, news articles in the previous 120 months (including month t) are used to estimate the sLDA model, and then articles in month t are used to compute topic weights in that month.

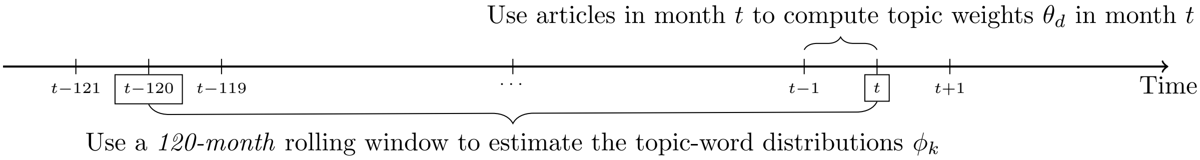


Figure 2. NYT Article Count and Length

This figure plots the time series of the monthly total count and the monthly average length of articles in the NYT. Article length is measured as the sum of unigrams (one-word terms), bigrams (two-word terms), and trigrams (three-word terms) of each article. Articles with limited content have been removed. The sample period is from January 1871 to October 2019.

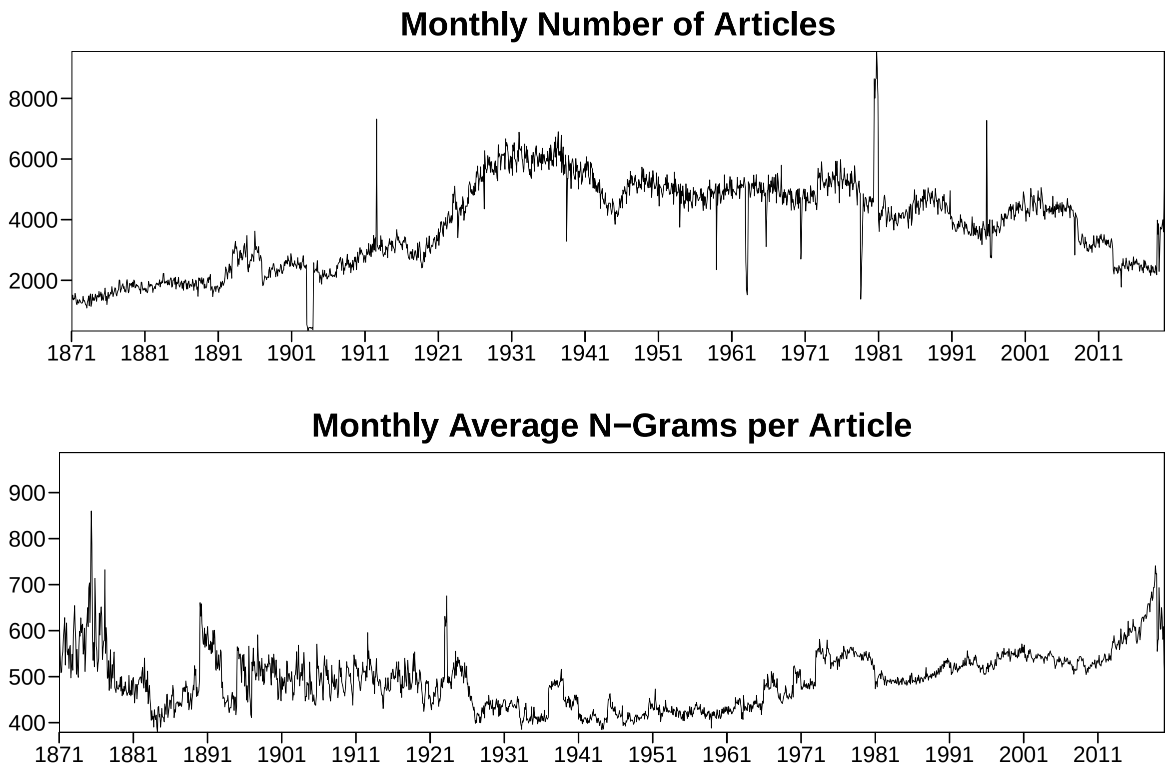
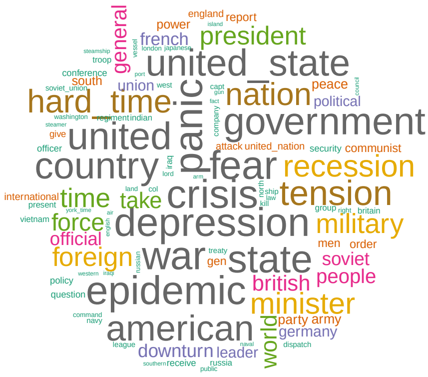


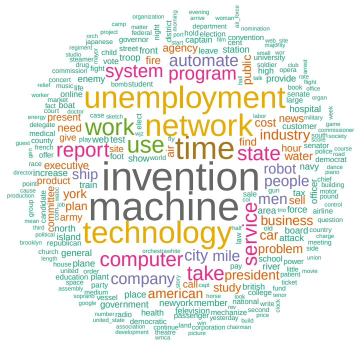
Figure 3. Narrative Contents

This figure plots the frequencies of ngrams per topic over time. Frequencies are constructed according to the sLDA model described in Section 2, and the size of each ngram indicates its frequency. The sample period is from January 1871 to October 2019.

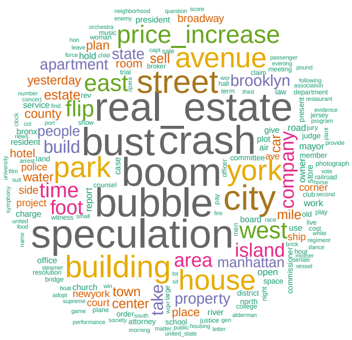
Panic



Technology Replacing Jobs



Real Estate Booms and Busts



Stock Market Bubbles

Boycotts and Evil Business

Figure 4. Time Series of Narrative Weights

This figure plots the time series of monthly topic weights constructed according to the sLDA model described in Section 2. The solid line represents the topic weight, and the dashed line represents the excess market return; both have been demeaned to improve visualization. The gray-shaded areas represent NBER-defined recessions. The sample period is from January 1871 to October 2019.

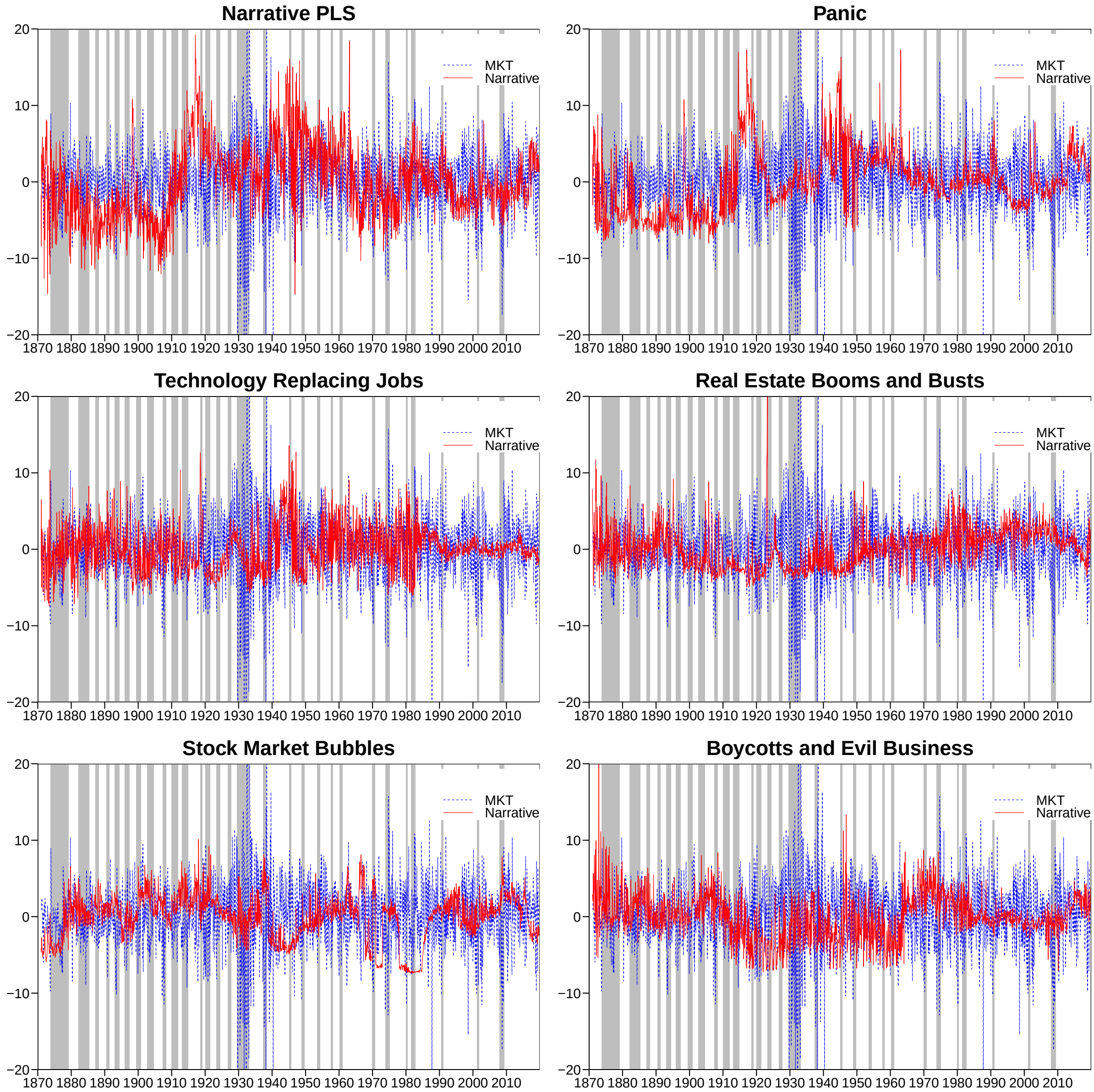


Figure 5. Articles Making the Biggest Contribution to Panic and Stock Spikes since 1990

This figure plots the 20 articles that have had the biggest contribution to 20 monthly heights of Panic and Stock since 1990. Topic weights are demeaned. The gray-shaded areas represent NBER-defined recessions. The sample period is from January 1990 to October 2019.

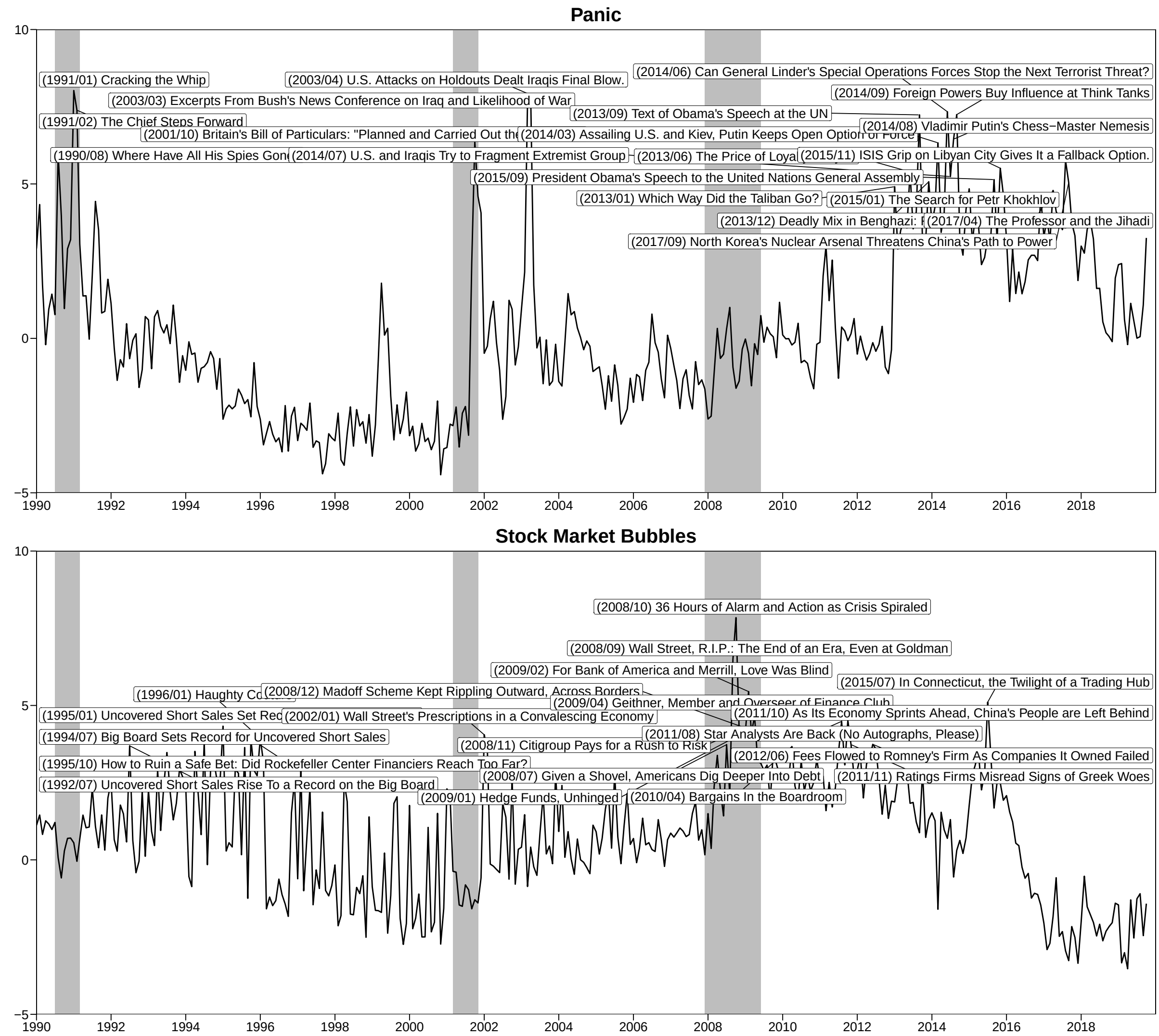


Figure 6. Cumulative In-Sample R^2 in One-Month Return Prediction

This figure plots the cumulative in-sample R^2 computed as

$$\left(\sum_{s=1}^{t}\left(R_s^e-\bar{R}^e\right)^2-\sum_{s=1}^{t}\left(R_s^e-\hat{R}_s^e\right)^2\right)/\sum_{s=1}^{T}\left(R_s^e-\bar{R}^e\right)^2,$$

where $\bar{R}^e$ is the sample mean of excess return and $\hat{R}_s^e$ is the fitted value from regression (2). The sample period is from January 1871 to October 2019.

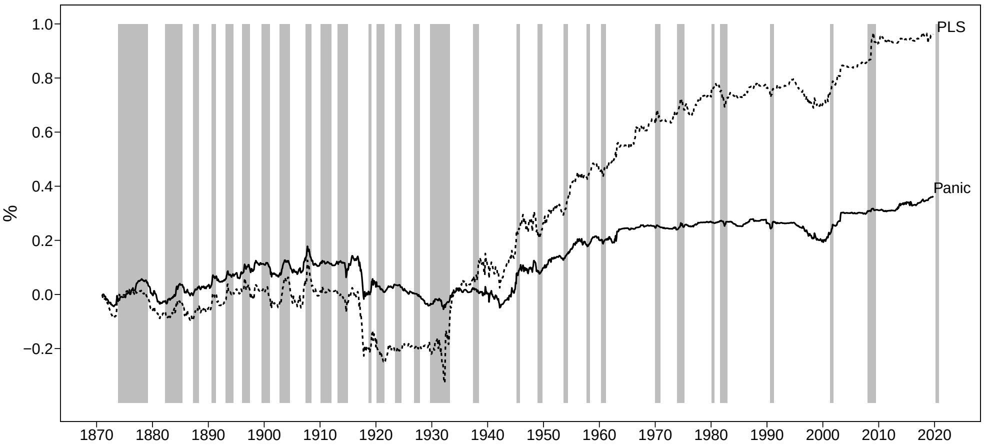


Figure 7. Cumulative Out-of-Sample R^2 in One-Month Return Prediction

This figure plots the cumulative out-of-sample R^2 computed as

$$\left(\sum_{s=1}^{t}\left(R_s^e-\bar{R}_s^e\right)^2-\sum_{s=1}^{t}\left(R_s^e-\hat{R}_s^e\right)^2\right)/\sum_{s=1}^{T}\left(R_s^e-\bar{R}_s^e\right)^2,$$

where $\bar{R}_s^e$ and $\hat{R}_s^e$ are, respectively, the historical mean and predicted value, estimated based on the preceding estimation window. The evaluation period is from January 1891 to October 2019.

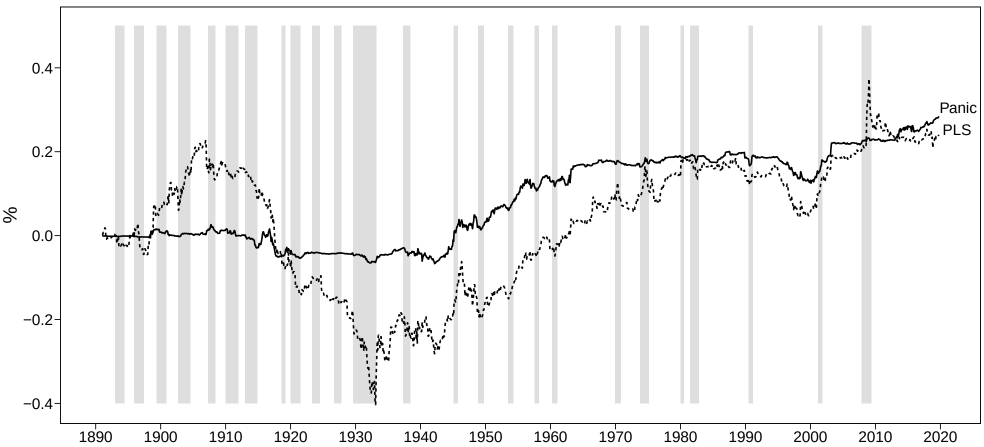


Table 1
Seed Words

This table lists the lemmatized seed words for each of the 10 narratives. The first column presents the full name of the narrative, and the second column reports the short name used in the paper.

Narrative	Short Name	Seed Words
Panic	Panic	bank failure, bank panic, bank run, crisis, depression, downturn, epidemic, fear, financial panic, great depression, great recession, hard time, pandemic, panic, recession, tension, war
Confidence	Confidence	autosuggestion, business confidence, confidence, consumer confidence, crowd psychology, suggestability
Savings	Saving	compassion, extravagance, family morale, frugal, frugality, modesty, moral, poverty, save, saving
Conspicuous Consumption	Consumption	american dream, conspicuous consumption, consumption, equal opportunity, equality, homeownership, luxury, patriotism, prosperity
Monetary Standard	Money	bimetallism, bitcoin, cryptocurrency, devaluation, gold, gold standard, inflation, monetary standard, money, silver
Technology Replacing Jobs	Tech	artificial intelligience, automate, computer, digital divide, electronic brain, internet, invention, labor save, labor save machine, machine, machine learn, mechanize, network, robot, technocracy, technological unemployment, technology, unemployment
Real Estate Booms and Busts	Real estate	boom, bubble, bust, crash, flip, flipper, home index, home ownership, home price, home purchase, house bubble, house index, house market, house price, land boom, land bubble, land price, price increase, real estate, real estate boom, real estate bubble, speculation
Stock Market Bubbles	Stock	advance market, bear, bearish, boom, boom and crash, bubble, bull, bull market, bullish, bust, crash, earnings per share, fall market, inflate market, margin, margin requirement, market boom, market bubble, market crash, price earn ratio, price increase, sell short, short sell, speculation, stock, stock crash, stock market boom, stock market bubble, stock market crash, stock market decline
Boycotts and Evil Business	Boycott	anger, boycott, community, evil business, excess profit, fair wage, moral, outrage, postpone purchase, profiteer, protest, strike, wage cut
Wage and Labor Unions	Wage	consumer price, cost of live, cost push, cost push inflation, high wage, increase wage, inflation, labor union, rise cost, wage, wage demand, wage lag, wage price, wage price spiral

Table 2
Summary Statistics

This table presents the summary statistics for the time series of 10 monthly topic weights constructed according to the sLDA model described in Section 2. Panel A reports the pairwise correlations with the topic weights based on the raw count of seed words; Panel B reports the first and second moments; Panel C reports the autocorrelations from the first to fourth orders; Panel D reports the loading on each topic in constructing a partial least squares (PLS) narrative index; and Panel E reports the correlations among topics. All numbers (except sample size) are expressed as percentages. The sample period is from January 1871 to October 2019.

	Panic	Confidence	Saving	Consumption	Money	Tech	Real_estate	Stock	Boycott	Wage	PLS
Panel A: Correlations with Raw Topic Counts											
Correlation	68.73 ***	2.25	17.64 ***	14.62 ***	6.38 ***	13.16 ***	26.23 ***	32.32 ***	12.45 ***	-2.37	60.81 ***
Panel B: Summary Statistic											
N	1784	1784	1784	1784	1784	1784	1784	1784	1784	1784	1784
Mean	11.92	9.15	10.72	10.45	8.59	9.02	7.54	8.15	8.49	8.05	32.67
SD	4.13	3.51	4.41	3.79	2.80	2.94	2.62	3.01	2.99	2.77	48.48
Q1	9.04	6.58	7.49	7.22	6.66	7.16	5.31	6.65	6.83	6.23	0.16
Median	11.69	8.72	10.50	10.49	8.21	8.78	7.43	8.57	8.44	7.78	29.23
Q3	14.29	11.59	13.90	13.48	10.37	10.66	9.29	10.04	10.36	9.40	62.86
Panel C: Autocorrelations											
AC(1)	78.19	6.44	57.52	15.41	60.59	30.77	50.79	82.12	51.73	6.96	65.25
AC(2)	75.58	6.85	56.78	12.85	57.27	30.90	48.05	79.07	47.09	3.81	62.37
AC(3)	72.50	6.95	55.61	14.88	55.38	29.37	42.81	77.00	46.30	8.94	61.29
AC(4)	67.91	8.18	54.08	11.46	51.60	27.76	37.58	73.84	41.67	7.11	56.23
Panel D: PLS Weights											
Weights	5.23	1.94	-0.37	0.92	-0.88	-2.98	-1.01	-2.50	-3.32	1.69	
Panel E: Correlations											
Panic		-10.06	-20.52	-10.89	-7.59	3.14	-20.48	-9.77	-30.30	-7.93	81.29
Confidence	-10.06		-7.85	-16.88	-15.64	-9.71	-4.84	-8.26	-9.02	-6.41	22.45
Saving	-20.52	-7.85		2.41	-29.45	-18.94	-23.79	-9.71	-8.45	-7.68	-7.29
Consumption	-10.89	-16.88	2.41		-9.52	-15.39	-18.55	-6.56	-11.69	-16.81	9.79
Money	-7.59	-15.64	-29.45	-9.52		-7.57	5.45	-8.09	5.46	-9.89	-16.15
Tech	3.14	-9.71	-18.94	-15.39	-7.57		-1.51	-14.16	-5.27	-6.12	-24.28
Real_estate	-20.48	-4.84	-23.79	-18.55	5.45	-1.51		-11.13	6.40	0.80	-24.47
Stock	-9.77	-8.26	-9.71	-6.56	-8.09	-14.16	-11.13		-18.02	-6.27	-23.11
Boycott	-30.30	-9.02	-8.45	-11.69	5.46	-5.27	6.40	-18.02		-7.64	-53.25
Wage	-7.93	-6.41	-7.68	-16.81	-9.89	-6.12	0.80	-6.27	-7.64		13.47
PLS	81.29	22.45	-7.29	9.79	-16.15	-24.28	-24.47	-23.11	-53.25	13.47	

Table 3
Predicting One-Month Market Returns

This table presents the results of the following predictive regression:

$$R^e_{t+1} = \alpha + \beta x_t + \epsilon_{t+1\rightarrow t+1},$$

where R^e_{t+1} is the excess market return over the next month, x_t is one of the narratives or the PLS index, and β, the coefficient of interest, measures the strength of predictability. Returns are expressed as annualized percentages, and the independent variable is standardized to zero mean and unit variance. Adjusted R^2 is expressed as a percentage, and t-statistics are computed with Newey and West (1987) standard errors. *p <0.1; **p <0.05; ***p <0.01.

	1871-2019	1871-1949	1950-2019	2000-2019
Panic (%)	3.44 ***	2.67 *	5.00 ***	10.85 ***
t-stat	(3.18)	(1.76)	(3.12)	(3.74)
R^2 (%)	0.31	0.07	0.90	4.22
Confidence (%)	1.50	0.65	2.45	1.59
t-stat	(1.21)	(0.35)	(1.50)	(0.49)
R^2 (%)	0.01	-0.10	0.12	-0.32
Saving (%)	-0.23	-0.34	1.95	-3.81
t-stat	(-0.17)	(-0.20)	(1.05)	(-0.93)
R^2 (%)	-0.05	-0.10	0.03	0.15
Consumption (%)	0.66	3.15	-1.21	-2.98
t-stat	(0.47)	(1.61)	(-0.75)	(-0.94)
R^2 (%)	-0.04	0.15	-0.06	-0.07
Money (%)	-0.85	-0.96	-1.59	1.03
t-stat	(-0.65)	(-0.48)	(-0.90)	(0.31)
R^2 (%)	-0.03	-0.08	-0.02	-0.38
Tech (%)	-2.75 *	-1.78	-4.99 ***	-2.50
t-stat	(-1.90)	(-0.88)	(-2.82)	(-0.82)
R^2 (%)	0.18	-0.03	0.90	-0.18
Real Estate (%)	-1.05	-0.46	-3.73 **	-7.00 *
t-stat	(-0.74)	(-0.25)	(-2.19)	(-1.83)
R^2 (%)	-0.02	-0.10	0.45	1.51
Stock (%)	-2.25	-3.12	-1.20	-3.98
t-stat	(-1.64)	(-1.53)	(-0.63)	(-0.86)
R^2 (%)	0.10	0.14	-0.06	0.20
Boycott (%)	-3.02 **	-2.90	-3.64 **	0.20
t-stat	(-2.25)	(-1.50)	(-2.27)	(0.07)
R^2 (%)	0.22	0.11	0.42	-0.42
Wage (%)	1.66	0.58	2.56	2.34
t-stat	(1.45)	(0.35)	(1.57)	(0.65)
R^2 (%)	0.03	-0.10	0.15	-0.21
PLS (%)	5.60 ***	4.50 **	7.89 ***	11.75 ***
t-stat	(4.36)	(2.45)	(4.78)	(3.13)
R^2 (%)	0.91	0.41	2.42	5.02

Table 4
Predicting Long-Horizon Market Returns

This table presents the results of the following predictive regression:

$$R^e_{t+1\rightarrow t+h} = \alpha + \beta x_t + \epsilon_{t+1\rightarrow t+h},$$

where $R^e_{t+1\rightarrow t+h}$ is the excess market return over the next h months, x_t is either Panic or the PLS index, and β, the coefficient of interest, measures the strength of predictability. Returns are expressed as annualized percentages, and the independent variable is standardized to zero mean and unit variance. Adjusted R^2 is expressed as a percentage, and t-statistics are computed with Newey and West (1987) standard errors using the corresponding h lags. *p <0.1; **p <0.05; ***p <0.01.

	Panic (%)	t-stat	R^2 (%)	PLS (%)	t-stat	R^2 (%)	N
Panel A: 1871-2019							
h = 1	3.44 ***	(3.18)	0.31	5.60 ***	(4.36)	0.91	1784
h = 3	2.62 ***	(2.71)	0.47	3.47 ***	(3.08)	0.86	1784
h = 6	2.64 ***	(2.62)	1.02	3.00 ***	(2.80)	1.33	1784
h = 12	2.49 **	(2.31)	1.53	2.71 **	(2.47)	1.83	1784
h = 24	1.83 *	(1.84)	1.45	2.36 **	(2.55)	2.45	1776
h = 36	1.92 *	(1.91)	2.16	2.73 ***	(2.88)	4.43	1764
Panel B: 1871-1949							
h = 1	2.67 *	(1.76)	0.07	4.50 **	(2.45)	0.41	948
h = 3	2.10	(1.51)	0.15	3.14 *	(1.91)	0.47	948
h = 6	2.22	(1.54)	0.52	2.74 *	(1.77)	0.84	948
h = 12	2.26	(1.50)	0.94	2.84 *	(1.80)	1.55	948
h = 24	1.48	(1.14)	0.71	2.41 *	(1.94)	2.07	948
h = 36	1.51	(1.20)	1.07	2.84 **	(2.28)	4.02	948
Panel C: 1950-2019							
h = 1	5.00 ***	(3.12)	0.90	7.89 ***	(4.78)	2.42	836
h = 3	3.58 ***	(2.78)	1.38	4.22 ***	(3.18)	1.96	836
h = 6	3.35 **	(2.55)	2.18	3.59 ***	(2.83)	2.52	836
h = 12	2.79 *	(1.84)	2.68	2.58 **	(1.98)	2.28	836
h = 24	2.25	(1.45)	2.92	2.34 *	(1.92)	3.16	828
h = 36	2.44	(1.57)	4.49	2.64 **	(2.17)	5.27	816
Panel D: 2000-2019							
h = 1	10.85 ***	(3.74)	4.22	11.75 ***	(3.13)	5.02	238
h = 3	8.01 ***	(4.10)	6.72	6.25 **	(2.44)	3.92	238
h = 6	6.02 ***	(3.15)	6.62	3.92 *	(1.85)	2.57	238
h = 12	5.80 **	(2.51)	11.14	2.18	(1.17)	1.21	238
h = 24	5.03 **	(2.24)	14.25	1.57	(0.98)	1.00	230
h = 36	4.89 ***	(2.72)	20.16	1.72	(1.41)	2.09	218

Table 5
Predicting Market Returns after Controlling for Economic Variables

This table presents the results of the following predictive regression (in Panel A):

$$R^e_{t+1} = \alpha + \gamma z_t + \epsilon_{t+1},$$

the following predictive regression (in Panel B):

$$R^e_{t+1} = \alpha + \beta x_t + \gamma z_t + \epsilon_{t+1},$$

where R^e_{t+1} is the excess market return over the next month, x_t is the narrative PLS index, and z_t is one of the 16 economic variables: 14 economic predictors from Goyal and Welch (2008), output gap from Cooper and Priestley (2009), and short interest from Rapach et al. (2016). The last row reports the results using the PLS index constructed from 16 economic variables. Returns are expressed as annualized percentages, and the independent variable is standardized to zero mean and unit variance. Adjusted R^2 is expressed as a percentage, and t-statistics are computed with Newey and West (1987) standard errors. *p <0.1; **p <0.05; ***p <0.01.

	Panel A: Univariate		Panel B: Bivariate			
Economic Predictor	$\gamma(\%)$	$R^2(\%)$	$\beta(\%)$	$\gamma(\%)$	$R^2(\%)$	Period
Dividend-price ratio (DP)	1.39	0.00	5.53 ***	0.33	0.85	187101-201910
Dividend yield (DY)	2.03	0.07	5.41 ***	0.96	0.88	187102-201910
Earnings-price ratio (EP)	2.46	0.13	5.31 ***	1.13	0.89	187101-201910
Dividend payout ratio (DE)	-1.00	-0.03	5.57 ***	-0.82	0.87	187101-201910
Stock variance (SVAR)	-0.08	-0.06	5.61 ***	-0.32	0.85	187101-201910
Book-to-market Ratio (BM)	5.19	0.58	7.33 ***	3.19	1.72	192103-201910
Net equity expansion (NTIS)	-4.11	0.31	8.50 ***	-4.25	1.94	192612-201910
Treasury bill rate (TBL)	-3.68 **	0.36	5.08 ***	-2.72 *	1.07	187101-201910
Long term bond yield (LTY)	-2.82	0.11	7.76 ***	-0.82	1.41	191901-201910
Long term bond return (LTR)	3.36 *	0.18	8.41 ***	3.57 *	1.78	192601-201910
Term spread (TMS)	-2.70	0.10	7.80 ***	-0.68	1.41	191901-201910
Default yield spread (DFY)	2.87	0.12	7.75 ***	2.00	1.50	191901-201910
Default return spread (DFR)	2.30	0.04	8.37 ***	2.45	1.62	192601-201910
Inflation (INFL)	-3.32	0.20	6.71 ***	-3.54	1.27	191302-201910
Output Gap (OG)	-3.39	0.20	7.63 ***	-2.46	1.53	191902-201910
Short Interest (SI)	-5.70 **	0.94	6.73 ***	-5.18 **	2.32	197301-201412
Economic PLS	4.40 *	0.48	6.43 **	2.77	1.63	197301-201412

Table 6
Predicting Market Returns after Controlling for Uncertainty and Sentiment Variables

This table presents the correlation between the narrative PLS index and each uncertainty and sentiment variable (in Panel A); the results of the following predictive regression (in Panel B):

$$R^e_{t+1} = \alpha + \gamma z_t + \epsilon_{t+1};$$

and the results of the following predictive regression (in Panel C):

$$R^e_{t+1} = \alpha + \beta x_t + \gamma z_t + \epsilon_{t+1},$$

where R^e_{t+1} is the excess market return over the next month, x_t is the narrative PLS index, and z_t is one of the uncertainty variables (financial and macro uncertainty indexes from Jurado et al. (2015), economic policy uncertainty index from Baker et al. (2016), disagreement index from Huang et al. (2020), implied volatility (VIX), and news implied volatility (NVIX) from Manela and Moreira (2017)), sentiment variables (news sentiment, investor sentiment from Baker and Wurgler (2006), aligned sentiment from Huang et al. (2015), and manager sentiment from Jiang et al. (2019)), or Shiller's confidence indexes: one-year confidence index and crash confidence index. The last row reports the results using the PLS index constructed from all uncertainty and sentiment variables. Returns are expressed as annualized percentages, and the independent variable is standardized to zero mean and unit variance. Adjusted R^2 is expressed as a percentage, and t-statistics are computed with Newey and West (1987) standard errors. *p <0.1; **p <0.05; ***p <0.01.

	Panel A: Correlations	Panel B: Univariate		Panel C: Bivariate			
Economic Predictor	Corr. with PLS (%)	γ(%)	R^2(%)	β(%)	γ(%)	R^2(%)	Period
Financial uncertainty	-10.34 ***	-5.75 **	1.15	6.97 ***	-5.03 *	2.88	196007-201910
Macro uncertainty	-3.48	-4.30	0.58	7.35 ***	-4.04	2.54	196007-201910
Economic policy uncertainty	21.42 ***	4.03	0.38	6.90 **	2.55	1.89	198501-201910
Implied volatility (VIX)	-12.74 **	0.40	-0.27	7.33 **	1.34	1.65	199001-201910
News implied volatility (NVIX)	2.33	0.03	-0.07	6.20 ***	-0.11	0.93	188907-201603
Disagreement	-1.76	-8.43 ***	2.43	6.03 ***	-8.32 ***	3.60	196912-201812
News sentiment	5.79 **	-0.52	-0.05	5.65 ***	-0.84	0.87	186612-201910
Investor sentiment (BW)	13.34 ***	-2.50	0.08	7.31 ***	-3.48	1.90	196507-201812
Investor sentiment (PLS)	7.21 *	-7.32 ***	1.86	7.41 ***	-7.86 ***	3.77	196507-201812
Manager sentiment	-17.53 **	-9.06 ***	3.32	10.18 **	-7.28 **	7.55	200301-201712
Shiller's one-year confidence index	-28.77 ***	-4.77	0.48	10.18 **	-1.84	3.96	200107-201910
Shiller's crash confidence index	-1.19	-2.07	-0.28	10.68 **	-1.95	3.99	200107-201910
Uncertainty PLS	-10.59	2.38	-0.39	12.58 ***	3.71	5.72	200301-201603

Table 7
Risk-Return Trade-Off

This table presents the results of the GARCH-M framework with the constant relative risk aversion specification (constant RRA):

$$R_{M,t+1} - R_{f,t+1} = \lambda_0 + \lambda_1 \times \sigma^2_{M,t} + \epsilon_{t+1},$$

and the time-varying RRA specification (varying RRA):

$$R_{M,t+1} - R_{f,t+1} = \lambda_0 + (\lambda_1 + \lambda_2 \times Panic_t) \times \sigma^2_{M,t} + \epsilon_{t+1},$$

in the *mean* equation. Panels A–D report the results with different specifications for the *volatility* equation, namely, GARCH (Bollerslev, 1986), IGARCH (Engle and Bollerslev, 1986), TGARCH (Zakoian, 1994), and EGARCH (Nelson, 1991):

$$\begin{aligned}
\text{GARCH}(1,1): \sigma^2_{M,t} &= \delta_0 + \delta_1 \epsilon^2_t + \delta_2 \sigma^2_{M,t-1},\\
\text{IGARCH}(1,1): \quad \sigma^2_{M,t} &= \delta_0 + \delta_1 \epsilon^2_t + (1-\delta_1)\sigma^2_{M,t-1},\\
\text{TGARCH}(1,1): \quad \sigma^2_{M,t} &= \delta_0 + \delta_1 \epsilon^2_t + \delta_3 D_t \epsilon^2_t + \delta_2 \sigma^2_{M,t-1},\\
\text{EGARCH}(1,1): \quad ln(\sigma^2_{M,t}) &= \delta_0 + \delta_1 \left(\frac{|\epsilon_t|}{\sigma_{M,t}}\right) - \delta_3 \left(\frac{\epsilon_t}{\sigma_{M,t}}\right) + \delta_2 ln(\sigma^2_{M,t-1}),
\end{aligned}$$

where D_t is an indicator equal to one when ϵ_t is negative and zero otherwise. The coefficient of interest λ_2, which measures the sensitivity of RRA to Panic, is in bold. The whole sample is January 1871 to October 2019.

	1871-2019				1871-1949				1950-2019			
	Constant RRA		Varying RRA		Constant RRA		Varying RRA		Constant RRA		Varying RRA	
	Coef.	*t*-stat	Coef.	*t*-stat	Coef.	*t*-stat	Coef.	*t*-stat	Coef.	*t*-stat	Coef.	*t*-stat
Panel A: GARCH												
λ_0	0.00	1.47	0.00	-0.82	0.00	0.37	0.00	-0.60	0.00	1.09	-0.01	-1.25
λ_1	2.17	2.62	0.43	0.41	2.20	2.42	1.24	1.10	2.58	1.32	-1.79	-0.76
λ_2			**1.58**	**2.93**			**0.94**	**1.53**			**3.60**	**3.12**
δ_0	0.00	3.56	0.00	3.60	0.00	2.67	0.00	2.69	0.00	2.47	0.00	2.48
δ_1	0.14	6.42	0.14	6.55	0.16	4.92	0.16	5.00	0.12	3.75	0.12	3.70
δ_2	0.82	31.85	0.82	32.74	0.81	20.95	0.81	21.36	0.83	24.50	0.83	23.66
Adj. R^2(%)	-0.38		0.17		-0.73		-0.44		0.10		1.23	
Panel B: IGARCH												
λ_0	0.00	2.17	0.00	-0.71	0.00	0.68	0.00	-0.54	0.00	1.95	0.00	-1.15
λ_1	1.69	2.84	0.13	0.17	1.81	2.41	0.91	0.96	1.88	1.80	-2.04	-1.32
λ_2			**1.53**	**3.68**			**0.94**	**1.95**			**3.15**	**3.24**
δ_0	0.00	3.60	0.00	3.62	0.00	2.57	0.00	2.58	0.00	2.93	0.00	2.85
δ_1	0.18	6.65	0.17	6.67	0.20	4.65	0.19	4.70	0.16	5.50	0.16	5.28
δ_2	0.82		0.83		0.80		0.81		0.84		0.84	
Adj. R^2(%)	-0.32		0.21		-0.61		-0.33		0.11		1.08	
Panel C: TGARCH												
λ_0	0.00	2.11	0.00	-1.16	0.00	0.43	0.00	-0.32	0.01	4.19	0.00	-0.89
λ_1	2.33	10.18	0.32	0.27	2.09	2.02	1.23	0.08	0.31	0.33	-3.05	-1.87
λ_2			**1.65**	**3.77**			**1.01**	**0.14**			**3.09**	**7.17**
δ_0	0.00	6.48	0.00	4.37	0.00	4.01	0.00	1.72	0.01	1.33	0.01	2.05
δ_1	0.14	11.79	0.14	7.18	0.14	8.72	0.14	0.64	0.12	4.52	0.12	5.29
δ_2	0.84	805.33	0.85	55.67	0.85	616.62	0.85	6.55	0.76	7.04	0.77	12.28
δ_3	0.26	3.38	0.27	3.25	0.20	2.29	0.21	1.96	0.76	1.71	0.63	2.11
Adj. R^2(%)	-0.26		0.14		-0.35		-0.50		-0.11		0.81	
Panel D: EGARCH												
λ_0	0.00	1.17	0.00	-0.74	0.00	-0.95	0.00	-3.84	0.01	1.62	0.00	-0.83
λ_1	2.88	43.43	0.53	1.30	2.81	16.15	1.38	0.93	0.89	0.44	-2.78	-1.41
λ_2			**1.61**	**2.31**			**0.94**	**4.38**			**3.17**	**3.74**
δ_0	-0.28	-43.00	-0.27	-0.76	-0.21	-23.38	-0.18	-2.16	-0.81	-2.77	-0.78	-1.09
δ_1	-0.06	-3.75	-0.06	-1.61	-0.05	-2.38	-0.05	-2.64	-0.14	-2.59	-0.12	-1.52
δ_2	0.96	65420.25	0.96	17.46	0.97	4136.73	0.97	72.89	0.88	19.74	0.88	7.99
δ_3	0.26	8.27	0.25	5.90	0.26	5.47	0.26	4.86	0.22	5.46	0.22	5.36
Adj. R^2(%)	-0.56		0.12		-1.17		-0.46		-0.13		0.89	

Table 8
Predicting Market Volatility

This table presents the results of the following predictive regression:

$$\sigma_{t+1\to t+h} = \alpha + \beta Panic_t + \delta' W_t + \epsilon_{t+1\to t+h},$$

where $\sigma_{t+1\to t+h}$ is the market volatility over the next h months and W_t is a set of controls. Panel A (B) reports the results for realized (implied) volatility. When σ_t is realized volatility, W_t includes two lags of realized volatility and two lags of negative market returns; when V_t is implied volatility (VIX), W_t includes two lags of VIX, two lags of realized volatility, and two lags of negative market returns. Volatility is expressed as annualized percentages, and the independent variable is standardized to zero mean and unit variance. Adjusted R^2 is expressed as a percentage, and t-statistics are computed with Newey and West (1987) standard errors using the corresponding $h + 2$ lags. *p <0.1; **p <0.05; ***p <0.01.

	Panic (%)	t-stat	R^2 (%)	N
Panel A: Realized Volatility				
1927-2019				
h = 1	-0.56 ***	(-3.98)	59.18	1112
h = 3	-0.69 ***	(-4.12)	52.67	1112
1950-2019				
h = 1	-0.59 ***	(-4.11)	47.05	836
h = 3	-0.72 ***	(-4.15)	39.30	836
2000-2019				
h = 1	-0.72 **	(-2.39)	59.09	238
h = 3	-1.04 ***	(-3.00)	45.43	238
Panel B: VIX (1990-2019)				
h = 1	-0.53 ***	(-3.93)	81.83	356
h = 3	-0.78 ***	(-4.05)	69.07	356

Table 9
Predicting Market Returns after Controlling for Real Events

This table presents the results of the following predictive regression:

$$R^e_{t+1} = \alpha + \beta \times Panic_t + \gamma^j \times D^j_t + \epsilon_{t+1}$$

where R^e_{t+1} is the excess market return over the next month, D^j_t is a dummy variable for event j equal to one if there is one event j in month t. Returns are expressed as annualized percentages, and Panic is standardized to zero mean and unit variance. Adjusted R^2 is expressed as a percentage, and t-statistics are computed with Newey and West (1987) standard errors. *p <0.1; **p <0.05; ***p <0.01.

	Recessions	Bank Failures	Wars	Disasters	Epidemic	All
Panel A: 1871-2019						
Panic	2.79 **	3.46 ***	3.21 ***	3.38 ***	3.46 ***	3.35 ***
	(2.50)	(3.20)	(2.85)	(3.13)	(3.20)	(3.13)
Event	-8.68 **	2.79	3.09	-6.92	17.14	-9.08 ***
	(-2.21)	(0.25)	(0.76)	(-1.08)	(1.52)	(-3.17)
R^2(%)	0.71	0.25	0.30	0.30	0.30	0.88
Panel B: 1871-1949						
Panic	1.74	2.77 *	1.79	2.58 *	2.67 *	2.66 *
	(1.08)	(1.83)	(1.07)	(1.70)	(1.76)	(1.76)
Event	-11.55 **	11.83	9.91	-8.39	19.68	-11.81 ***
	(-2.38)	(0.99)	(1.39)	(-1.18)	(1.47)	(-2.90)
R^2(%)	0.76	0.01	0.28	0.02	0.05	0.85
Panel C: 1950-2019						
Panic	5.04 ***	4.95 ***	5.10 ***	4.93 ***	5.00 ***	5.15 ***
	(3.12)	(3.09)	(3.16)	(3.07)	(3.12)	(3.18)
Event	-4.47	-31.47 *	-2.63	-5.76	8.68	-6.15
	(-0.63)	(-1.74)	(-0.60)	(-0.59)	(0.49)	(-1.56)
R^2(%)	0.87	0.93	0.83	0.84	0.79	1.13

Table 10
Out-of-Sample R^2

This table reports the out-of-sample R^2_{OS} statistic (Campbell and Thompson, 2008) in predicting the monthly excess market return using economic narratives. Panels A, B, and C report the results from an OLS regression, the PLS index, and random forest, respectively. All the out-of-sample forecasts are estimated recursively using the data available in the expanding estimation window. All numbers are expressed as percentages. The evaluation period begins in January 1891, and the whole sample is from January 1871 to October 2019. $*p<0.1$; $**p<0.05$; $***p<0.01$ (based on the Clark and West (2007) MSFE-adjusted statistic).

	1891-2019	1891-1949	1950-2019	2000-2019
Panel A: OLS				
Dividend-price ratio (DP)	-0.39	-0.48	-0.25	0.05
Dividend yield (DY)	-0.34	-0.15	-0.64	0.04
Earnings-price ratio (EP)	-0.05	0.07	-0.26	-0.35
Dividend payout ratio (DE)	-0.65	-0.84	-0.33	-1.06
Stock variance (SVAR)	-1.67	-2.19	-0.79	-0.86
Treasury bill rate (TBL)	0.07 **	-0.04	0.26 **	0.45
Panic	0.28 ***	0.05	0.68 ***	1.41 ***
Confidence	-0.04	-0.12	0.10	0.03
Saving	-0.10	-0.10	-0.11	-0.11
Consumption	-0.18	0.01	-0.50	-0.19
Money	-0.14	-0.19	-0.05	-0.12
Tech	0.11 *	-0.09	0.46 **	0.12
Real_estate	-0.06	-0.11	0.02	0.22
Stock	-0.00	0.06	-0.11	0.34
Boycott	0.15 **	0.03	0.36 **	-0.17
Wage	0.01	-0.08	0.15	0.10
Panel B: PLS				
Economic	-0.62	-0.76	-0.38	-0.55
Narrative	0.24 **	-0.26	1.07 ***	1.71 *
Panel C: Random Forrest				
Economic	-11.41	-8.96	-15.54	-6.83
Narrative	-2.17	-1.53	-3.25	0.22

Table 11
Asset Allocation Results

This table reports the annualized certainty equivalent returns (utility) gains as percentages and the annualized monthly Sharpe ratio for a mean-variance trading strategy. The strategy uses 6 economic predictors or 10 narratives to make return forecasts compared to using historical mean returns as the forecasts. Panels A, B, and C report the results using OLS, PLS, and random forest regressions, respectively. The last row reports the annualized monthly Sharp ratio of the S&P 500 index. All of the out-of-sample forecasts are estimated recursively using the data available in the expanding estimation window. The evaluation period begins in January 1891, and the whole sample is from January 1871 to October 2019. *p <0.1; **p <0.05; ***p <0.01 (based on the test statistics in DeMiguel et al. (2009)).

	Utility Gain (%)				Sharpe Ratio			
	1891-2019	1891-1949	1950-2019	2000-2019	1891-2019	1891-1949	1950-2019	2000-2019
Panel A: OLS								
Dividend-price ratio (DP)	-0.35	0.04	-0.70	0.24	0.40	0.30	0.48	0.21
Dividend yield (DY)	-0.53	-0.06	-0.94	1.02	0.39	0.29	0.50	0.30
Earnings-price ratio (EP)	0.39	0.09	0.64	3.88 *	0.46	0.31	0.56	0.65 ***
Dividend payout ratio (DE)	0.32	0.78	-0.08	-0.10	0.44	0.37	0.49	0.26
Stock variance (SVAR)	-0.58	-0.64	-0.53	-0.44	0.37	0.23	0.46	0.23
Treasury bill rate (TBL)	1.43 **	1.08 *	1.72 *	1.93 *	0.52 **	0.39 *	0.62 **	0.40 **
Panic	0.56 **	0.19	0.87 ***	1.70 ***	0.46 **	0.32	0.55 ***	0.36 ***
Confidence	-0.02	-0.14	0.08	0.27	0.41	0.28	0.50	0.26
Saving	-0.22	-0.45	-0.03	-0.01	0.40	0.25	0.49	0.25
Consumption	-0.58	-0.25	-0.88	-0.12	0.37	0.28	0.44	0.23
Money	-0.03	-0.19	0.10	0.03	0.42	0.28	0.50	0.23
Tech	0.40	-0.16	0.89 **	0.13	0.45	0.28	0.56 **	0.26
Real Estate	-0.14	-0.35	0.04	0.35	0.41	0.26	0.50	0.26
Stock	0.42	-0.02	0.79	1.72	0.45	0.30	0.55	0.35
Boycott	0.43	0.50	0.37	-0.33	0.45	0.36	0.52	0.23
Wage	0.09	-0.14	0.28	0.17	0.42	0.28	0.51	0.26
Panel B: PLS								
Economic	0.34	0.54	0.16	3.08	0.45	0.35	0.53	0.65 **
Narrative	0.92 *	-0.01	1.71 **	3.03 *	0.49 *	0.31	0.63 **	0.46 *
Panel C: Random Forrest								
Economic	0.56	1.90	-0.60	1.83	0.45	0.45 *	0.46	0.37
Narrative	1.13 *	2.10 **	0.30	4.45 **	0.50	0.47 **	0.52	0.55 **
Buy and Hold					0.42	0.32	0.55	0.35

Table 12
Subperiod R^2

This table reports the R^2 statistic as a percentage computed over different subperiods: expansion (exp) versus recession (rec) and high sentiment versus low sentiment. Expansions and recessions are based on National Bureau of Economic Research (NBER) business cycles. A month is classified as high (low) sentiment if the Baker and Wurgler (2006) investor sentiment level in the previous month is above (below) the median value for the sample. Panel A reports the results for the in-sample analysis, and the full sample period is January 1971 to October 2019. Panel B reports the results for the out-of-sample analysis with an expanding estimation window, and the evaluation period begins in January 1891.

	R^2	R^2_{exp}	R^2_{rec}	R^2_{high}	R^2_{low}
Panel A: In Sample					
Panic	0.31	0.11	0.67	0.08	0.69
PLS	0.91	0.90	1.03	0.55	2.09
Panel B: Out of Sample					
Panic	0.28	0.51	0.00	0.06	0.64
PLS	0.24	0.81	-0.47	0.04	1.18

Table 13
Predicting Long-Horizon Market Returns with Narratives from the WSJ

This table presents the results of the following predictive regression:

$$R^e_{t+1\to t+h} = \alpha + \beta x_t + \epsilon_{t+1\to t+h},$$

where $R^e_{t+1\to t+h}$ is the excess market return over the next h months, x_t is Panic, Stock, or the PLS index, and β, the coefficient of interest, measures the strength of predictability. Returns are expressed as annualized percentages, and the independent variable is standardized to zero mean and unit variance. Adjusted R^2 is expressed as a percentage, and t-statistics are computed with Newey and West (1987) standard errors using the corresponding h lags. The sample period is from January 2000 to October 2019. *p <0.1; **p <0.05; ***p <0.01.

	Panic (%)	t-stat	R^2 (%)	Stock (%)	t-stat	R^2 (%)	PLS (%)	t-stat	R^2 (%)	N
h = 1	7.24 **	(2.13)	1.65	-8.63 ***	(-2.64)	2.51	9.01 ***	(2.63)	2.78	238
h = 3	5.34 **	(1.96)	2.75	-5.57 **	(-2.53)	3.03	4.00 *	(1.83)	1.35	238
h = 6	5.10 **	(2.39)	4.63	-4.53 **	(-1.96)	3.56	4.67 ***	(2.62)	3.81	238
h = 12	5.46 ***	(3.03)	9.84	-4.60 **	(-2.12)	6.86	4.33 **	(2.44)	6.02	238
h = 24	5.13 ***	(2.65)	14.79	-3.46	(-1.61)	6.49	4.00 **	(2.33)	8.84	230
h = 36	5.17 ***	(3.90)	22.56	-2.36	(-1.33)	4.35	3.94 ***	(2.99)	12.94	218

Table 14
Out-of-Sample R^2: WSJ Stock versus NYT Panic

This table reports the out-of-sample R^2_{OS} statistic (Campbell and Thompson, 2008) in predicting excess market returns over the next h months using economic narratives. All of the out-of-sample forecasts are estimated recursively using the data available in the expanding estimation window. Reported are the mean R^2_{OS} and mean p-value based on the Clark and West (2007) MSFE-adjusted statistic across all different initial estimation windows, ranging from 60 months to 180 months with an increment of 12 months. The whole sample is January 2000 to October 2019.

		h=1	h=3	h=6	h=12
WSJ Stock	R^2_{OS} (%)	2.721	3.368	1.044	3.479
	p-value	0.050	0.143	0.283	0.205
NYT Panic	R^2_{OS} (%)	1.120	8.144	10.879	10.865
	p-value	0.007	0.003	0.010	0.008

Internet Appendix
Economic Narratives and Market Outcomes: A Semi-supervised Topic Modeling Approach

A Text Processing Steps

Before carrying out text cleaning, we first remove articles with limited contents. Specifically, we remove articles whose title contains the following terms: *a day s weddings, advertising amp marketing, advertising and marketing news, advertising news, advertising news and notes, amusements this evening, apartment leases, apartment rentals, army and navy, army orders and assignments, around the garden, arrival of buyers, arrivals at the hotels, arrivals in the city, art, arts, assets and liabilities in central reserve cities, at the hotels, at the movies, auction, auctions, bankrupt notices, bankruptcy notices, bankruptcy proceedings, bankruptcy sales, baseball, basketball, bond notes, boston stock market, briefs debt issues, bronx mortgages filed, bronx properties sold new dealings in improved and unimproved holdings, bronx transfers, building plans filed, business leases, business troubles, butter and egg market, by cable, calendar, calendars, chicago live stock, chicago produce markets, chicago quotations, churches and ministers home and foreign events, classified advertisements, coast guard orders, coming events, commercial affairs, commercial leases, commodity cash prices, corporate changes, corporate reports, country produce markets, court of appeals, cricket, crossword puzzle, current issues and yields, customs patent appeals court, deals and discounts, death list of a day, decisions supreme court chambers, delaware charters, diner s journal, dining out style desk, dinner menu for tonight, directory to dining, display ad, dividend meetings, dividend news, dividends announced, dividends announced dividend meetings today, dividends declared, events today, executive changes, federal courts, fire, food notes, football, foreign markets, garden, garden q amp a house amp home style desk, general markets, going out guide, government maturities, guide for buyers, guide schedule, home improvement, homes that sold for around, in this issue, incoming steamships, index to classified advertisements, insiders stockholdings, legal advertisement, legal advertisements, legal notice, legal notices, live stock in chicago, live stock markets, livestock in chicago, locally dressed meats, long is-*

land guide, mail ships, major league baseball, major league leaders, manhattan mortgages, manhattan transfers, marine and aviation reports, marine corps orders, marine intelligence, market averages, markets by telegraph, minor leagues, missing persons, money and credit, money and credit bullion, money and exchange, movements of naval vessels, movies critic s choice television, movies this week television, music notes, n f l matchups week, naval orders, new buildings and alterations, new incorporations, new jersey guide, new york cattle market, new york charters, news of the advertising and marketing fields, news of wood field and stream, notes of insurance interests, notes of the stage, notes of various interests, notes on fashion, obituary, ocean travel, ocean travelers, of local origin, off the menu, off the menu dining in dining out style desk, offerings and yields of municipal bonds, on television, on the market real estate desk, other company reports, other corporate reports, out of town exchanges, outgoing steamships, passengers arrived, passengers sailed, philadelphia prices, pickups and putouts, police department, pop and jazz guide review, post and paddock, pro transactions, proceedings, produce markets, programs of the week, puns and anagrams, q amp a, q amp a question, quotation of the day, raceway entries, real estate transactions, real estate transfers, realty financing, recorded leases, referees notices, reports on ski conditions, reports on skiing conditions, reserve corps orders, residential resales list, residential sales around the region list, residential sales list, restaurants review, results plus, round about the garden, sales list, saturday news quiz question, schedule, schedules, securities at auction, shipping and foreign mails, shipping and mails, shipping and the mails, shipping mails, ski, soccer, social activities in new york and elsewhere, social notes, spare times schedule, sport, sports, statistical summary, stock exchange bid and asked, stock exchange news, stock quote, surrogate notices, the beauty quest, the boston market, the building department list of plans filed, the chicago market, the civil service, the cotton markets, the guide schedule, the proceedings in albany, the real estate market recorded real estate transfers, the record of accidents, the state of trade, the united service, the united service army, theatrical gossip, theatrical notes, toronto stock exchange, transactions list, transfers in the bronx, transfers recorded, travel advisory, treasure chest, utility earnings, weather, westchester guide, what s on tonight schedule, wills for probate, wood field and stream, yachts reported, year maturities are.

We further remove articles whose title exactly matches the following phrases: *accounts, amusements, arrived, bank notices, births, bridge, brooklyn, business records, chess, commodities markets, commodity prices, deaths, died, dividends, domestic markets, economic indicators, engagements, estates appraised, executives, federal reserve statement, finan-*

cial notes, fires, foreign exchange, foreign ports, highs and lows, in the real estate field, insurance, market averages, money, municipal loans, nuggets, proposal, proposals, q amp a travel desk, railroads, real estate notes, recorded mortgages, retail store sales, scouting, steamboats, sugar coffee cocoa, the london market, the pop life, the real estate market, the standings, transactions, treasury statement, utility reports. For example, an article whose title reads "Accounts" will be removed while one whose title is "North American segment accounts for 70% of total revenues." will not be removed.

These terms and phrases are obtained by manually checking the most frequent title patterns showing up in the news archive. We acknowledge that these lists are not comprehensive and articles whose content is limited are still present in the data.

Next, we conduct the following text cleaning steps:

[1] Remove articles with fewer than 100 content words. We consider content words as those <u>outside</u> of the expanded stop word list of 3,346 words developed by Professor Matthew L. Jockers. This list is available at `https://www.matthewjockers.net/macroanalysisbook/expanded-stopwords-list/`. We append this list with full and abbreviated day and month names (e.g. Monday, Mon, November, Nov, etc.).

[2] Turn all words into lower case and remove unicode code points, html tags, hashtags, urls, one-letter words and words containing three or more repeating letters.

[3] Lemmatize texts using part-of-speech tags. Part-of-speech tagging and lemmatization are conducted using the nltk library in Python.

[4] Tokenize texts into unigrams, bigrams, and trigrams within sentence punctuation boundaries. In natural language processing, "tokenize" means breaking document into words or "tokens." "Unigram" refers to a one-word token, "bigram" a two-word token, and "trigram" a three-word token. Collectively, "ngram"refers to an n-word token. To create sensible ngrams, it is important to retain punctuation before tokenization.

[5] Remove unigrams of fewer than three letters and ngrams containing stop words or numbers. For example, under within-punctuation boundary tokenization, the sentence "Under current favorable conditions, revenue of firm A will double next year." is converted into the following unigrams [current, favorable, condition, revenue, firm, double, year], bigrams [current_favorable, favorable_condition], and trigrams [current_favorable_condition] where all stop words and words of less than three characters have been removed.

[6] Each month t, with news articles over the past ten years up to and including month

t, we create a document-frequency matrix where each row is a document (article), each column is a token, and each entry is the count of the token in that document. To mitigate the impact of outliers on document-topic distribution, we remove tokens appearing in fewer than 0.2% and tokens appearing in more than 90% of all documents over the past ten years.

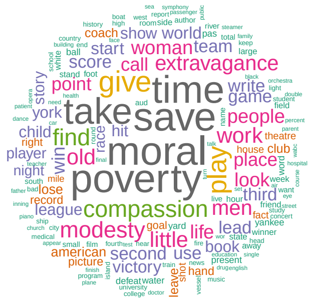
B Additional Figures and Tables for Economic Narratives from the NYT

Figure B1. Narrative Contents from the NYT

This figure plots the over-time frequencies of ngrams per each topic. The size of each ngram indicates its frequency. The sample period is from January 1871 to October 2019.

Confidence

Savings



Conspicuous Consumption

Monetary Standard

Wage and Labor Unions

Figure B2. Time Series of Narrative Weights from the NYT

This figure plots the time series of monthly topic weights constructed according to the sLDA model as described in Section 2. The solid line is topic weight while the dashed line is excess market return; both have been demeaned for ease of visualization. The shades indicate NBER-dated recessions. The sample period is from January 1871 to October 2019.

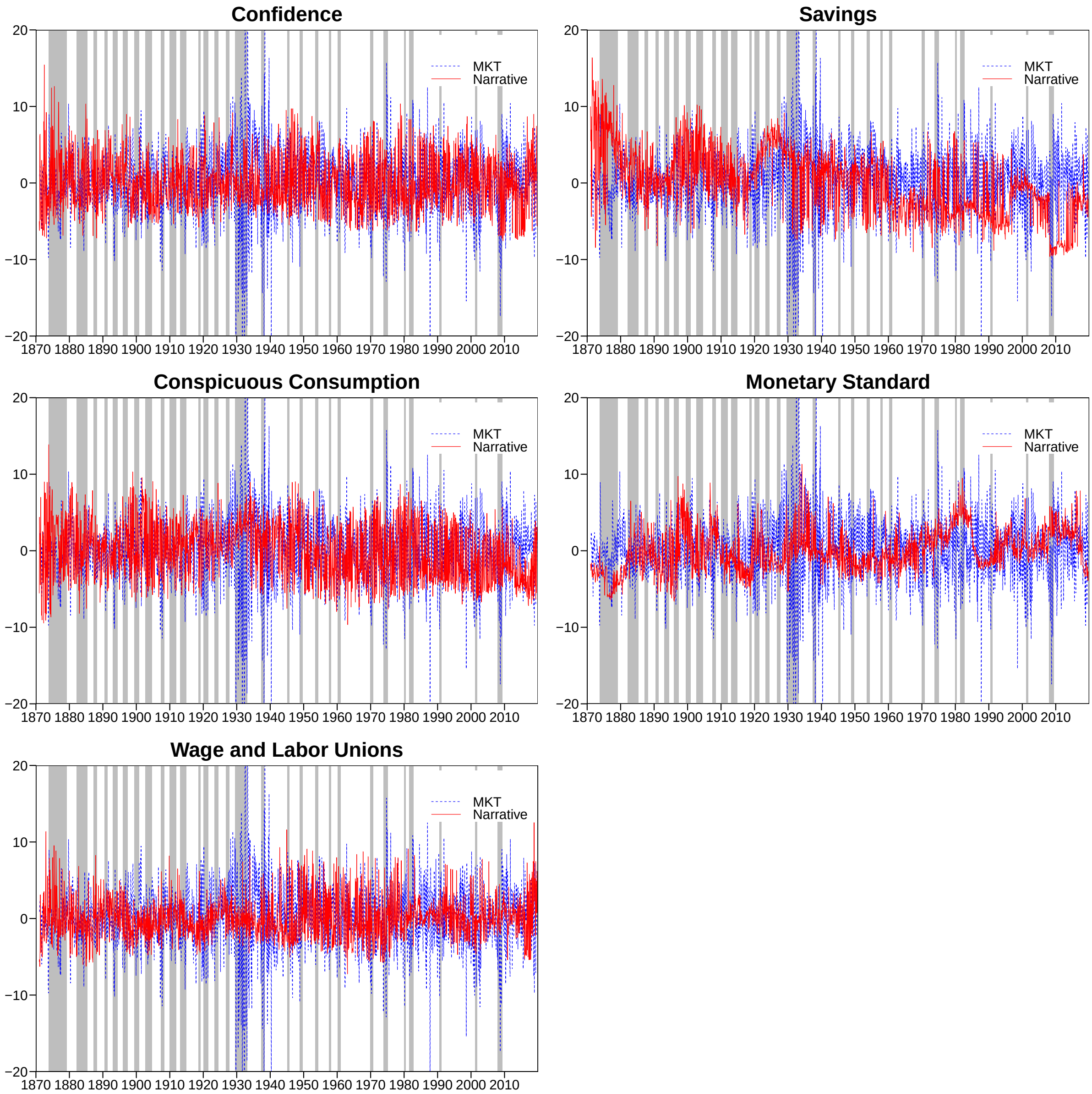


Figure B3. Historical Events

This figure plots the time series of monthly historical U.S. events. The horizal gray bar indicates there is at least one such labeled event in that month. Recession dates are from NBER while other event dates are from Global Financial Data. The sample period is from January 1871 to October 2019.

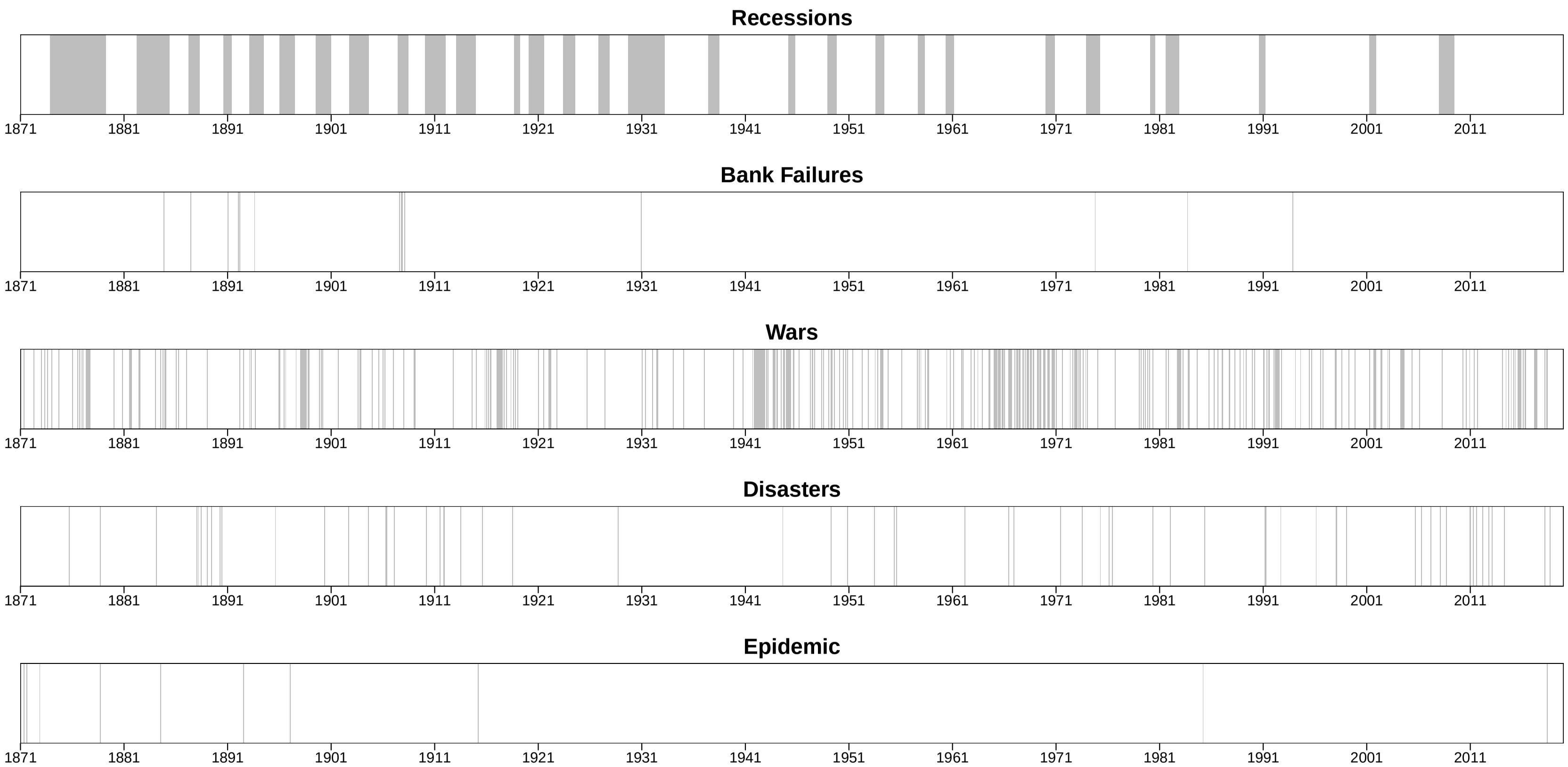


Table B1
Data Screening

This table reports the number of NYT articles after each cleaning step. The whole sample is from January 1871 to October 2019.

Screening Steps	Number of Articles (Millions)
Original Sample	14.73
After dropping articles whose title indicates limited content	13.41
After further dropping articles having fewer than 100 content words	6.89

Table B2
Out-Of-Sample R^2: 600-Month Rolling Window Estimation

This table reports the out-of-sample R^2_{OS} statistic (Campbell and Thompson, 2008) in predicting the monthly excess market return using the economic narratives. Panel A, B, and C reports results using OLS, PLS, and Random Forrest, respectively. All of the out-of-sample forecasts are estimated recursively using data available in the 600-month rolling estimation window. All numbers are in percentages. ***, **, and * indicate 1%, 5%, and 10% significance of the Clark and West (2007) MSFE-adj statistic. The evaluation period begins in January 1891 and the whole sample is from January 1871 to October 2019.

	1927-2019	1927-1949	1950-2019	2000-2019
Panel A: OLS				
Dividend-price ratio (DP)	-0.20	-0.34	-0.02 **	0.99 *
Dividend yield (DY)	-0.06 **	0.04	-0.20 **	1.06 *
Earnings-price ratio (EP)	-0.06	0.22	-0.42	0.15
Dividend payout ratio (DE)	-0.78	-0.56	-1.07	-1.24
Stock variance (SVAR)	-1.38	-2.64	0.24	1.96
Treasury bill rate (TBL)	-0.43	-1.08	0.40 **	-0.61
Panic	0.33 ***	0.09	0.63 ***	2.59 ***
Confidence	-0.12	-0.19	-0.02	-0.03
Saving	-0.07	-0.09	-0.06	-1.01
Consumption	-0.11	-0.10	-0.12	-0.01
Money	-0.03	-0.05	0.01	-0.01
Tech	0.06	-0.21	0.41 **	-0.11
Real_estate	0.09	-0.09	0.31 *	1.15 **
Stock	-0.05	0.12	-0.27	-0.05
Boycott	0.31 ***	0.23 *	0.40 **	-0.32
Wage	0.01	-0.06	0.09 *	-0.20
Panel B: PLS				
Economic	-0.25	-0.11	-0.42	0.05
Narrative	0.30 ***	-0.08	0.77 ***	0.54
Panel C: Random Forrest				
Economic	-10.79	-10.71	-10.90 *	-5.61
Narrative	-4.21	-4.90	-3.33 *	-0.48

Table B3
Predicting Returns of Characteristics Portfolios

This table presents the results of the following predictive regression:

$$R^e_{i,t+1} = \alpha_i + \beta_i x_t + \epsilon_{i,t+1}, \quad i = 1, \ldots, 40$$

where $R^e_{i,t+1}$ is the excess return on portfolio i over the next month, x_t is either Panic or the PLS index, and β_i, the coefficient of interest, measures the strength of predictability. Returns are expressed as annualized percentages, and the independent variable is standardized to zero mean and unit variance. Adjusted R^2 is expressed as a percentage, and t-statistics are computed with Newey and West (1987) standard errors. The sample period is January 1927 to October 2019. *p <0.1; **p <0.05; ***p <0.01.

	Panic (%)	t-stat	R^2 (%)	PLS (%)	t-stat	R^2 (%)
Panel A: Industry Portfolios						
Nondurable	2.36 *	(1.76)	0.10	5.17 ***	(3.45)	0.80
Durable	5.21 **	(2.42)	0.23	10.70 ***	(4.32)	1.27
Manufacture	2.65	(1.55)	0.04	7.54 ***	(3.60)	0.93
Energy	2.69	(1.48)	0.05	6.70 ***	(3.26)	0.75
Technology	4.15 *	(1.90)	0.14	8.62 ***	(3.70)	0.90
Telecom	0.74	(0.55)	-0.07	3.46 **	(2.41)	0.30
Shop	3.82 **	(2.24)	0.21	7.91 ***	(4.35)	1.21
Health	2.57	(1.53)	0.06	5.30 ***	(2.81)	0.54
Utility	1.80	(1.10)	-0.02	6.11 ***	(3.44)	0.77
Other	4.91 ***	(2.72)	0.32	9.55 ***	(4.18)	1.46
Panel B: Size Portfolios						
Small	9.27 ***	(3.00)	0.53	15.44 ***	(3.81)	1.62
2	7.10 ***	(2.84)	0.38	12.13 ***	(3.79)	1.27
3	6.65 ***	(2.99)	0.40	11.47 ***	(4.01)	1.38
4	5.65 ***	(2.70)	0.32	10.75 ***	(4.10)	1.38
5	5.30 ***	(2.68)	0.31	10.16 ***	(4.18)	1.38
6	4.75 **	(2.49)	0.25	9.86 ***	(4.23)	1.40
7	4.92 ***	(2.61)	0.32	9.52 ***	(4.31)	1.45
8	4.24 **	(2.45)	0.25	8.88 ***	(4.25)	1.38
9	3.64 **	(2.24)	0.19	7.75 ***	(3.99)	1.17
Large	3.11 **	(2.16)	0.18	6.83 ***	(4.17)	1.19
Panel C: Book-to-market Portfolios						
Growth	2.71 *	(1.65)	0.07	6.46 ***	(3.52)	0.82
2	3.49 **	(2.23)	0.21	7.53 ***	(4.41)	1.32
3	2.74 *	(1.81)	0.09	6.90 ***	(4.02)	1.07
4	2.46	(1.54)	0.03	6.62 ***	(3.29)	0.80
5	3.72 **	(2.42)	0.21	8.20 ***	(4.36)	1.38
6	2.94 *	(1.81)	0.08	7.05 ***	(3.38)	0.86
7	4.10 **	(2.35)	0.20	8.93 ***	(3.95)	1.27
8	4.53 **	(2.47)	0.23	9.90 ***	(4.12)	1.43
9	5.49 ***	(2.66)	0.27	11.34 ***	(4.10)	1.45
Value	8.05 ***	(3.00)	0.45	14.05 ***	(4.06)	1.57
Panel D: Momentum Portfolios						
Losers	7.73 ***	(2.75)	0.35	11.84 ***	(3.35)	0.94
2	4.14 *	(1.86)	0.10	9.36 ***	(3.22)	0.86
3	3.74 **	(1.97)	0.11	7.65 ***	(3.17)	0.76
4	3.18 *	(1.88)	0.09	7.42 ***	(3.52)	0.87
5	3.24 **	(2.00)	0.12	6.89 ***	(3.34)	0.86
6	3.65 **	(2.30)	0.19	7.88 ***	(4.08)	1.22
7	3.37 **	(2.14)	0.18	7.64 ***	(4.28)	1.29
8	3.21 **	(2.07)	0.17	7.48 ***	(4.31)	1.31
9	4.81 ***	(2.89)	0.43	9.12 ***	(5.04)	1.80
Winners	5.19 **	(2.56)	0.37	8.90 ***	(4.34)	1.25

Table B4
Predicting Daily Market Returns

This table presents the results of the following predictive regression:

$$R^e_{t+1\rightarrow t+h} = \alpha + \beta x_t + \delta' W_t + \epsilon_{t+1\rightarrow t+h},$$

where $R^e_{t+1\rightarrow t+h}$ is the excess market return over the next h periods; x_t is either Panic or the PLS index; W_t is vector of controls including five lags of excess market return, five lags of squared excess return, five lags of sentiment, and weekday indicators; and β, the coefficient of interest, measures the strength of predictability. Returns are in annualized percentages and the independent variable is standardized to zero mean and unit variance. Adjusted R^2 is expressed as a percentage, and t-statistics are computed with Newey and West (1987) standard errors using the corresponding $h + 5$ lags. The sample period is January 1927 to October 2019. *p <0.1; **p <0.05; ***p <0.01.

	Panic (%)	t-stat	R^2 (%)	PLS (%)	t-stat	R^2 (%)	N
Panel A: 1927-2019							
h = 1	4.23 **	(2.36)	0.79	6.41 ***	(3.54)	0.82	24306
h = 5	3.96 ***	(2.78)	0.15	4.72 ***	(3.31)	0.19	24306
Panel B: 1950-2019							
h = 1	3.62 **	(2.20)	0.61	4.39 **	(2.57)	0.62	17550
h = 5	4.21 ***	(3.33)	0.36	3.93 ***	(2.98)	0.34	17550
Panel C: 2000-2019							
h = 1	8.91 **	(2.23)	1.31	10.06 ***	(2.59)	1.35	4981
h = 5	8.75 ***	(3.31)	1.51	5.69 **	(2.15)	1.29	4981

C Economic Narratives from the WSJ

Before extracting the ten narratives from the WSJ articles, we also conduct text-processing steps. Similar to the procedure applied to the NYT articles, we remove articles with limited content indicated first by the pattern of the section they belong to if the section label is available and then by the pattern of their title. These section and title patterns are constructed by manually examining the articles and are available upon request. We then follow the procedure described in Appendix A to clean the texts and convert them into ngrams.

Figure C1. WSJ Article Count and Length from the WSJ

This figure plots the time series of monthly total count and monthly average length of articles in the WSJ. Article length is measured as the sum of unigrams (one-word terms), bigrams (two-word terms), and trigrams (three-word terms) of each article. Articles with limited contents have been removed. The sample period is from December 1899 to October 2019.

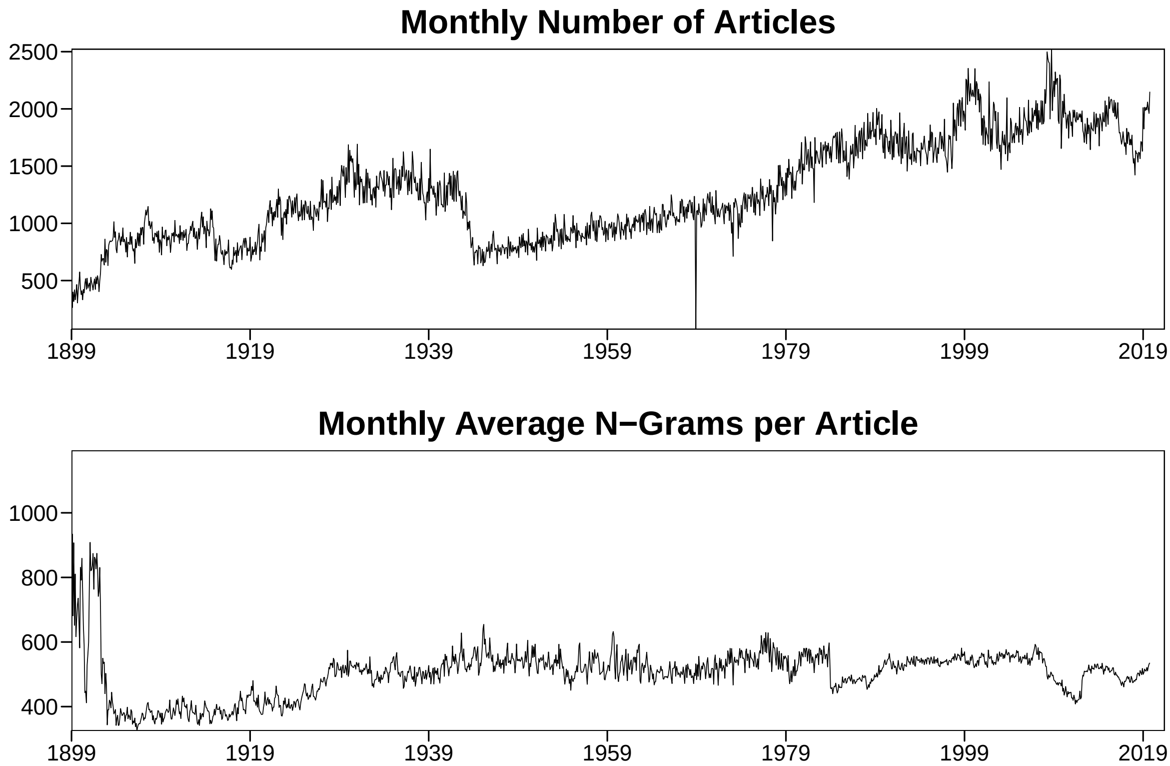
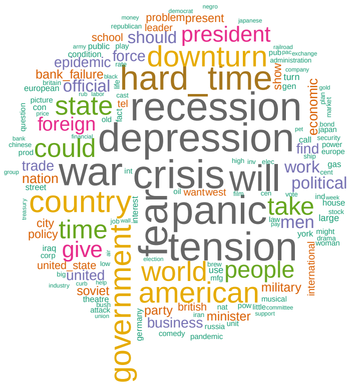


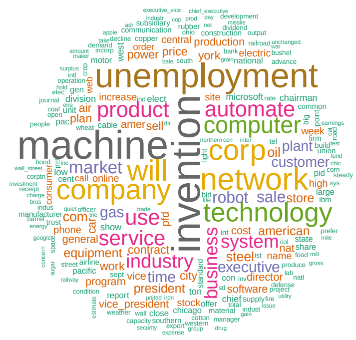
Figure C2. Narrative Contents from the WSJ

This figure plots the over-time frequencies of ngrams per each topic constructed according to the sLDA model. The size of each ngram indicates its frequency. The sample period is from December 1899 to October 2019.

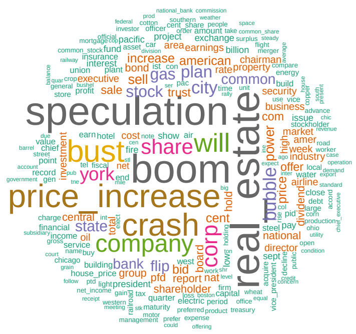
Panic



Technology Replacing Jobs



Real Estate Booms and Busts



Stock Market Bubbles

Boycotts and Evil Business

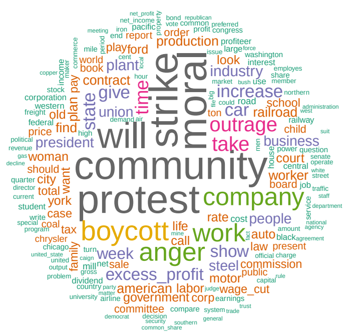
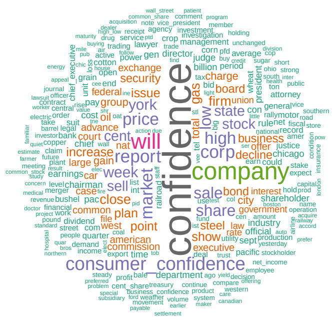


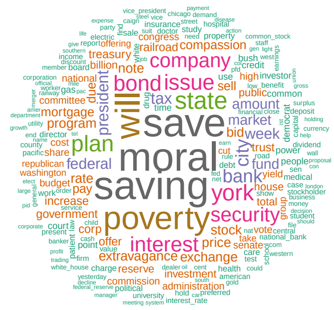
Figure C3. Narrative Contents from the WSJ

This figure plots the over-time frequencies of ngrams per each topic. The size of each ngram indicates its frequency. The sample period is from December 1899 to October 2019.

Confidence



Savings



Conspicuous Consumption

Monetary Standard

Wage and Labor Unions

Figure C4. Time Series of Narrative Weights from the WSJ

This figure plots the time series of monthly topic weights. The solid line is topic weight while the dashed line is excess market return; both have been demeaned for ease of visualization. The shades indicate NBER-dated recessions. The sample period is from December 1899 to October 2019.

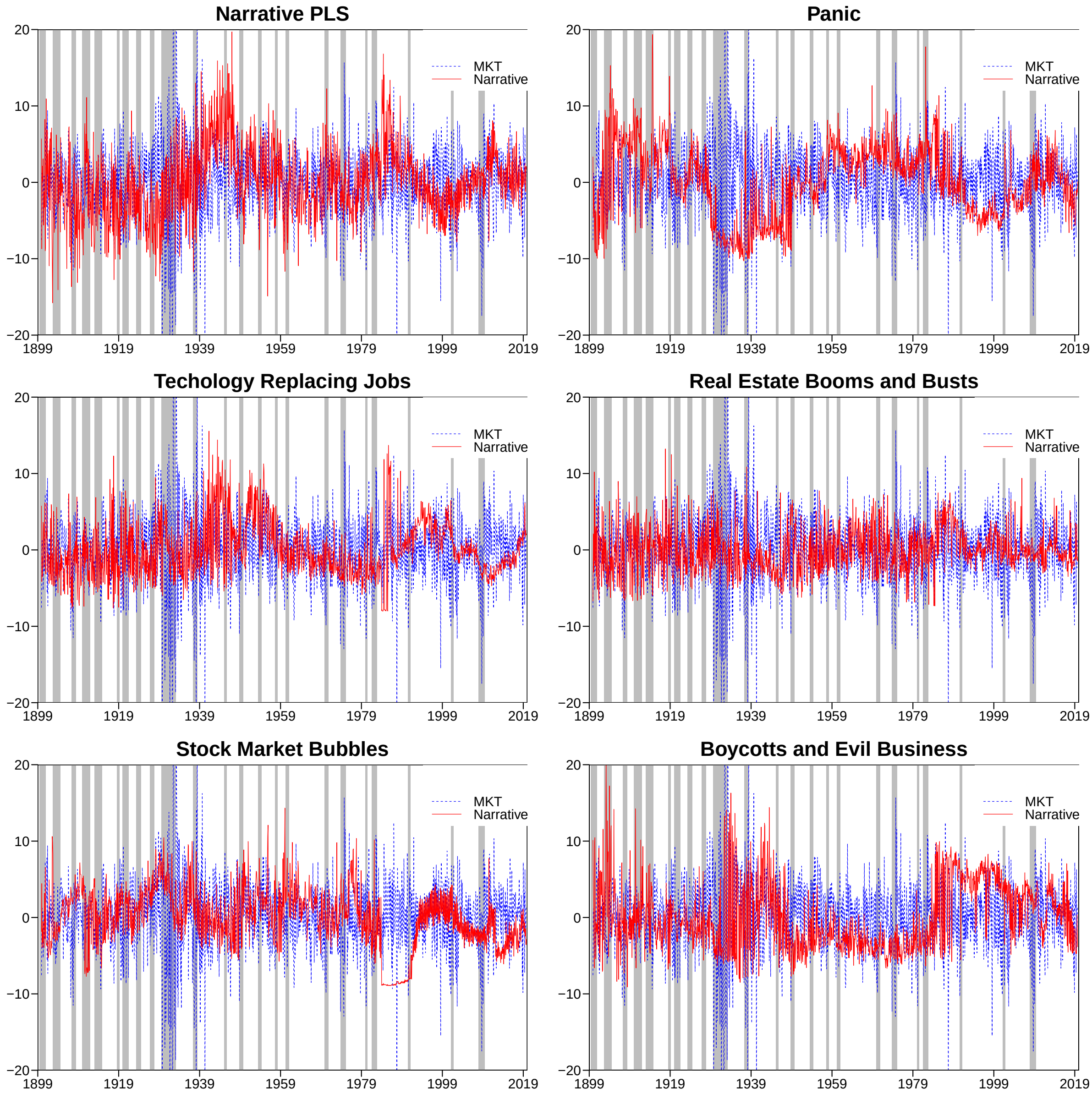


Figure C5. Time Series of Narrative Weights from the WSJ

This figure plots the time series of monthly topic weights. The solid line is topic weight while the dashed line is excess market return; both have been demeaned for ease of visualization. The shades indicate NBER-dated recessions. The sample period is from December 1899 to October 2019.

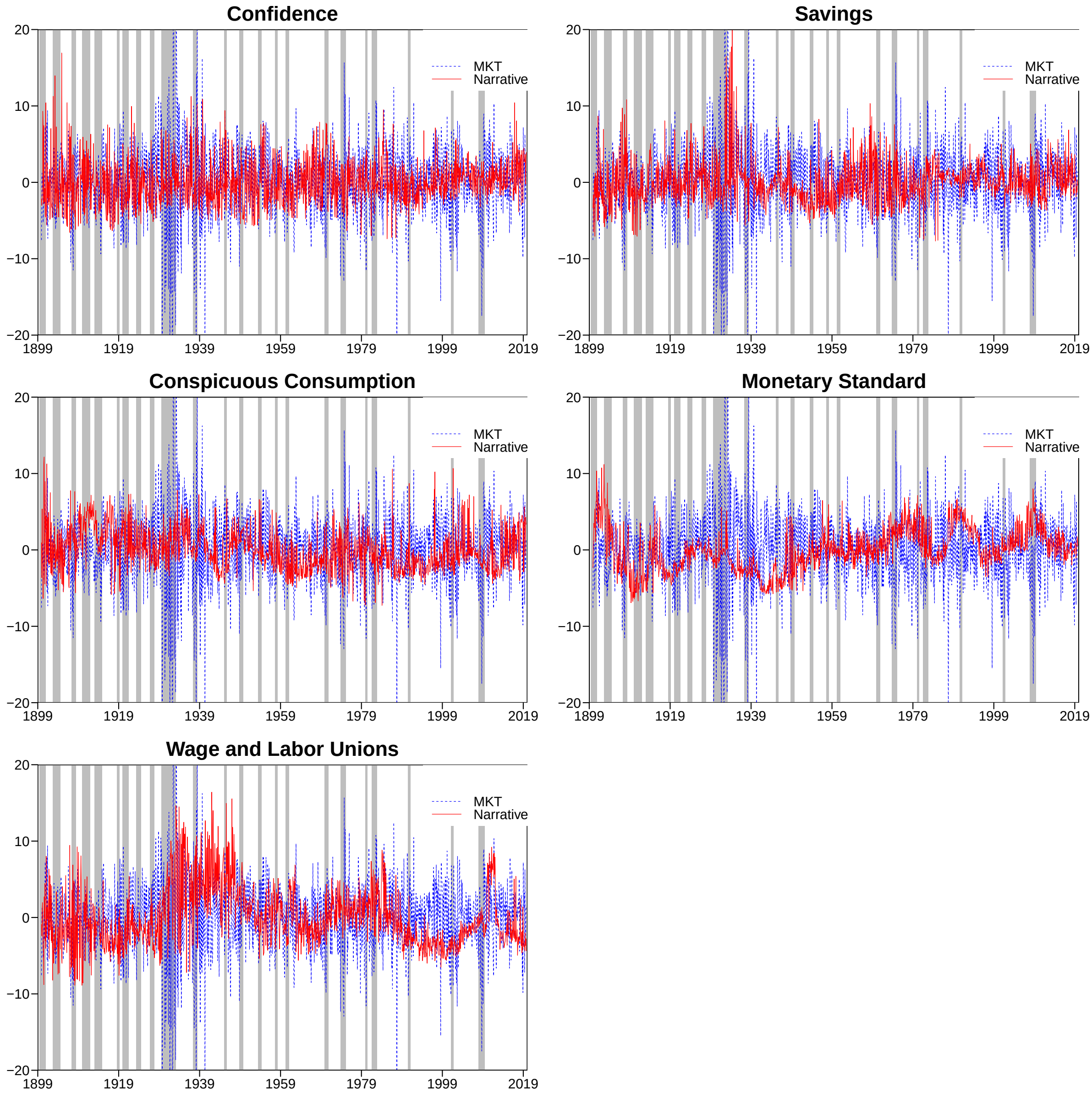


Table C1
Summary Statistics of Narratives from the WSJ

This table presents the summary statistic of the time series of ten monthly topic weights constructed according to the sLDA model as described in Section 2. Panel A reports the pairwise correlations with the topic weights from NYT; Panel B reports the first and second moments; Panel C reports the autocorrelations from first- to fourth-order; and Panel D reports the loading on each topic in constructing a partial least square (PLS) narrative index. All numbers (except sample size) are in percentages. The sample period is from December 1899 to October 2019.

	Panic	Confidence	Saving	Consumption	Money	Tech	Real_estate	Stock	Boycott	Wage	PLS
Panel A: Correlations with NYT Topic Weights											
Correlation	7.64 ***	2.70	-4.52 *	13.96 ***	16.06 ***	8.11 ***	6.40 **	11.34 ***	-9.84 ***	-3.08	14.14 ***
Panel B: Summary Statistic											
N	1439	1439	1439	1439	1439	1439	1439	1439	1439	1439	1439
Mean	12.65	7.92	8.29	7.91	7.94	8.85	7.94	9.51	10.68	10.25	12.00
SD	4.73	2.98	2.79	2.76	2.81	3.67	2.94	3.89	4.40	3.95	46.94
Q1	9.33	6.00	6.57	5.92	6.13	6.38	5.96	7.19	7.23	7.47	-17.94
Median	12.84	7.68	8.08	7.55	7.76	8.24	7.68	9.60	9.68	9.41	10.46
Q3	16.20	9.63	9.67	9.63	9.60	10.73	9.69	12.08	13.84	12.39	37.95
Panel C: Autocorrelations											
AC(1)	72.11	4.39	10.31	29.70	74.52	42.42	17.52	69.06	47.55	52.72	42.45
AC(2)	68.52	-0.22	15.12	29.77	68.46	38.67	17.83	62.03	40.88	51.79	39.17
AC(3)	66.49	4.48	18.26	35.57	67.44	37.74	17.50	63.46	42.44	51.30	40.59
AC(4)	65.83	3.07	9.30	25.67	67.46	42.73	15.22	60.13	44.24	47.41	37.66
Panel D: PLS Weights											
Weights	0.51	2.42	-0.17	-1.83	-0.70	2.74	-2.98	-4.09	-0.42	3.82	

Table C2
Predicting One-Month Market Returns with Narratives from the WSJ

This table presents the results from the following predictive regression:

$$R^e_{t+1} = \alpha + \beta x_t + \epsilon_{t+1\to t+1},$$

where R^e_{t+1} is the excess market return over the next month, x_t is one of the narratives or the PLS index, and β, the coefficient of interest, measures the strength of predictability. Returns are expressed as annualized percentages, and the independent variable is standardized to zero mean and unit variance. Adjusted R^2 is expressed as a percentage, and t-statistics are computed with Newey and West (1987) standard errors. The sample period is from December 1899 to October 2019. *p <0.1; **p <0.05; ***p <0.01.

	1899-2019	1899-1949	1950-2019	2000-2019
Panic (%)	0.33	0.23	0.31	7.24 **
t-stat	(0.15)	(0.07)	(0.19)	(2.13)
R^2 (%)	-0.07	-0.17	-0.12	1.65
Confidence (%)	2.50 *	2.74	2.29	-1.82
t-stat	(1.72)	(1.14)	(1.28)	(-0.58)
R^2 (%)	0.10	-0.03	0.09	-0.29
Saving (%)	-0.18	0.06	-0.41	1.52
t-stat	(-0.11)	(0.02)	(-0.25)	(0.51)
R^2 (%)	-0.07	-0.17	-0.11	-0.33
Consumption (%)	-2.04	-5.17 **	0.54	-0.53
t-stat	(-1.45)	(-2.08)	(0.32)	(-0.18)
R^2 (%)	0.04	0.33	-0.11	-0.41
Money (%)	-0.78	0.98	-2.98	-1.78
t-stat	(-0.49)	(0.38)	(-1.54)	(-0.35)
R^2 (%)	-0.05	-0.15	0.24	-0.30
Tech (%)	2.31	-0.33	4.59 ***	-1.98
t-stat	(1.49)	(-0.12)	(2.88)	(-0.53)
R^2 (%)	0.07	-0.16	0.74	-0.27
Real Estate (%)	-3.13 *	-6.65 **	-0.11	-3.17
t-stat	(-1.90)	(-2.43)	(-0.06)	(-1.02)
R^2 (%)	0.20	0.65	-0.12	-0.03
Stock (%)	-3.24 *	-3.11	-3.34 *	-8.63 ***
t-stat	(-1.92)	(-1.05)	(-1.88)	(-2.64)
R^2 (%)	0.21	0.01	0.33	2.51
Boycott (%)	-0.29	0.47	-0.90	-4.42
t-stat	(-0.13)	(0.11)	(-0.50)	(-1.58)
R^2 (%)	-0.07	-0.16	-0.09	0.35
Wage (%)	2.99	5.75	0.20	4.22
t-stat	(1.27)	(1.43)	(0.11)	(1.15)
R^2 (%)	0.17	0.44	-0.12	0.28
PLS (%)	6.57 ***	7.90 **	5.41 ***	9.01 ***
t-stat	(3.80)	(2.56)	(3.30)	(2.63)
R^2 (%)	1.10	0.98	1.07	2.78

D Topic Weights Constructed by Raw Counts of Seed Words

In this section, we conduct a robustness check for the main empirical results in the paper. Specifically, we investigate whether the sLDA model adds any economic insight beyond a simple count of seed words in the news.

While the majority of finance papers that employ textual analysis rely on simple counts of words from a predefined dictionary (for reviews, see Loughran and McDonald (2016) and Loughran and McDonald (2020)), recent studies have exploited statistical unsupervised topic modeling to extract thematic contents from textual data (e.g., Dyer et al. (2017), Choudhury et al. (2019), Brown et al. (2020), and Bybee et al. (2021)). This paper blends the two branches by employing a semisupervised model in which we inject seed words into the topic model to extract desired contents. Hence, a natural question is whether the sLDA model reveals any additional information beyond a simple count of those seed words in the news. To answer this question, we construct topic weights by simply counting the occurrences of seed words and scale them by the total number of ngrams in the article.

Table D1 reports the summary statistics for these topic weights. Panic is still the most frequently mentioned and most volatile topic with a monthly mean of 0.15% and standard deviation of 0.09%. Panic has a first-order autocorrelation of 96%, which is much higher than the percentage (78%) obtained via the sLDA one. To remain consistent with the sLDA model, we also construct the PLS index from all topics. Once again, the PLS index heavily loads on Panic and strongly correlates with this topic with a correlation coefficient of 98%.

To investigate whether manually constructed topics have the same market implications as the sLDA topics, we first use them to predict the monthly market returns in sample. Table D2 shows that in general, both Panic and the PLS index can strongly positively predict market excess returns one-month ahead, consistent with the sLDA results. One notable difference is that, while the manually constructed topics have stronger predictability before 2000 than do the sLDA-based topics, over the past 20 years, the manually constructed Panic index yields a much weaker predicting power over the next month than does the sLDA counterpart. The other manually constructed topics, similar to the sLDA topics, do not display any consistent predictability pattern.

In Table D3, we find that the manually constructed PLS index is not significant after controlling for the economic PLS index, and, in Table D4, the manually counted narrative index loses its significance when controlling for other uncertainty variables. Notably, the manually counted index is more correlated with other existing uncertainty indexes. For example, its correlation with the EPU index is 47%.

The in-sample predictability results can be biased if the predictors are highly consistent, which is the case for the manually counted Panic and PLS index. Hence, in Table D5, we report the out-of-sample R^2 computed with the frequency-based narratives. Unsurprisingly, over the whole evaluation period of 1891–2019, the raw Panic index produces a much lower R^2_{OS} than the sLDA one: 0.08% versus 0.28%. The sLDA one continues to outperform in each subperiod. Similarly, the manually constructed narratives via PLS greatly underperform their sLDA counterparts across all samples.

In sum, topic weights constructed with simple seed word counts yield monthly in-sample prediction results in line with the sLDA ones but substantially underperform in terms of out-of-sample predictability. Moreover, the frequency-based narrative index does not contain additional economic insights beyond the well-known economic and uncertainty predictors. These results indicate that the limited set of seed words fails to capture the whole universe of terms belonging to the same topic, and, hence, we need a statistical way to uncover and cluster them.

Table D1
Summary Statistics
Topic Weights Constructed by Raw Counts of Seed Words

This table presents the summary statistic of the time series of ten monthly topic weights from January 1871 to October 2019 constructed by raw counts of seed words. Panel A reports the first and second moments; Panel B reports the autocorrelations from first- to fourth-order; Panel C reports the loading on each topic in constructing a partial least square (PLS) narrative index; and Panel D report the correlations among topics. All numbers (except sample size) are in percentages.

	Panic	Confidence	Saving	Consumption	Money	Tech	Real_estate	Stock	Boycott	Wage	PLS
Panel A: Summary Statistic											
N	1784	1784	1784	1784	1784	1784	1784	1784	1784	1784	1784
Mean	0.15	0.01	0.05	0.02	0.12	0.06	0.03	0.15	0.10	0.03	49.62
SD	0.09	0.00	0.01	0.00	0.05	0.04	0.01	0.03	0.03	0.02	77.91
Q1	0.10	0.01	0.04	0.01	0.09	0.02	0.02	0.13	0.08	0.02	6.33
Median	0.13	0.01	0.05	0.02	0.12	0.04	0.03	0.14	0.10	0.02	32.16
Q3	0.16	0.01	0.05	0.02	0.14	0.09	0.04	0.16	0.12	0.03	62.16
Panel B: Autocorrelations											
AC(1)	95.68	67.69	74.78	74.01	87.80	97.70	85.19	83.69	72.27	84.77	95.88
AC(2)	92.36	58.48	68.65	67.70	79.90	97.12	82.84	77.94	57.33	76.97	92.62
AC(3)	89.94	51.69	65.10	63.06	74.57	96.65	80.57	74.89	50.30	72.00	90.30
AC(4)	88.49	48.30	60.31	61.97	69.83	96.13	80.62	72.81	46.73	68.64	88.89
Panel C: PLS Weights											
Weights	0.11	0.00	0.00	0.00	-0.01	0.02	0.00	-0.02	0.01	0.01	
Panel D: Correlations											
Panic		-0.31	2.55	21.41	-39.22	3.55	-12.94	-21.34	11.23	27.03	98.64
Confidence	-0.31		39.54	44.69	44.00	-24.15	-16.35	6.00	-27.63	-24.52	-7.23
Saving	2.55	39.54		30.89	38.59	-7.18	-10.41	4.88	-12.47	-19.04	-3.98
Consumption	21.41	44.69	30.89		22.79	-29.62	-9.86	8.40	-19.15	-10.00	14.59
Money	-39.22	44.00	38.59	22.79		-17.47	-1.80	13.95	-28.44	-22.60	-49.81
Tech	3.55	-24.15	-7.18	-29.62	-17.47		69.17	14.48	30.71	7.45	5.67
Real_estate	-12.94	-16.35	-10.41	-9.86	-1.80	69.17		42.03	19.11	-3.19	-15.46
Stock	-21.34	6.00	4.88	8.40	13.95	14.48	42.03		-3.12	-15.18	-32.35
Boycott	11.23	-27.63	-12.47	-19.15	-28.44	30.71	19.11	-3.12		45.36	14.18
Wage	27.03	-24.52	-19.04	-10.00	-22.60	7.45	-3.19	-15.18	45.36		29.25
PLS	98.64	-7.23	-3.98	14.59	-49.81	5.67	-15.46	-32.35	14.18	29.25	

Table D2
Predicting One-Month Market Returns: Raw Topic Counts

This table presents the results of the following predictive regression:

$$R^e_{t+1} = \alpha + \beta x_t + \epsilon_{t+1\to t+1},$$

where R^e_{t+1} is the excess market return over the next month, x_t is one of the narratives or the PLS index constructed by raw counts of seed words, and β is the coefficient of interest which measures the strength of predictability. Returns are in annualized percentages and the independent variable is standardized to zero mean and unit variance. Adjusted R^2 is in percentage and t-stat is computed with the Newey and West (1987) standard errors. ***, **, and * indicate significant at 1%, 5%, and 10% respectively.

	1871-2019	1871-1949	1950-2019	2000-2019
Panic (%)	3.49 ***	3.88 **	5.39 ***	8.04 *
t-stat	(2.90)	(2.41)	(2.90)	(1.91)
R^2 (%)	0.32	0.27	1.07	2.12
Confidence (%)	0.98	1.98	1.13	-7.28 *
t-stat	(0.44)	(0.59)	(0.53)	(-1.74)
R^2 (%)	-0.03	-0.01	-0.07	1.66
Saving (%)	0.21	0.96	0.16	1.72
t-stat	(0.13)	(0.38)	(0.09)	(0.46)
R^2 (%)	-0.05	-0.08	-0.12	-0.31
Consumption (%)	1.82	3.98 *	1.02	2.24
t-stat	(1.13)	(1.74)	(0.59)	(0.61)
R^2 (%)	0.05	0.29	-0.08	-0.23
Money (%)	-0.57	0.18	-1.00	-0.37
t-stat	(-0.36)	(0.08)	(-0.52)	(-0.08)
R^2 (%)	-0.05	-0.10	-0.08	-0.42
Tech (%)	1.22	-0.30	0.71	0.36
t-stat	(1.04)	(-0.11)	(0.43)	(0.10)
R^2 (%)	-0.01	-0.10	-0.10	-0.42
Real Estate (%)	0.62	0.08	-0.13	-1.07
t-stat	(0.45)	(0.03)	(-0.06)	(-0.21)
R^2 (%)	-0.04	-0.11	-0.12	-0.38
Stock (%)	-1.88	-2.13	-1.41	-7.00
t-stat	(-1.28)	(-1.04)	(-0.73)	(-1.56)
R^2 (%)	0.05	0.01	-0.04	1.51
Boycott (%)	1.15	1.01	0.22	7.29 ***
t-stat	(0.86)	(0.56)	(0.13)	(2.96)
R^2 (%)	-0.02	-0.08	-0.12	1.67
Wage (%)	2.13	3.34	-0.21	2.72
t-stat	(1.50)	(1.64)	(-0.10)	(0.93)
R^2 (%)	0.08	0.18	-0.12	-0.13
PLS (%)	3.48 ***	3.69 **	5.49 ***	9.19 ***
t-stat	(2.88)	(2.25)	(3.31)	(2.83)
R^2 (%)	0.32	0.24	1.11	2.91

Table D3
Predicting Market Returns after Controlling for Economic Variables
Topic Weights Constructed by Raw Counts of Seed Words

This table presents the results of the following predictive regression

$$R_{t+1}^e = \alpha + \gamma z_t + \epsilon_{t+1}$$

in Panel A, and the following predictive regression

$$R_{t+1}^e = \alpha + \beta x_t + \gamma z_t + \epsilon_{t+1}$$

in Panel B, where R_{t+1}^e is the excess market return over the next month, x_t is the narrative PLS index constructed by raw counts of seed words, and z_t is one of the 16 economic variables: 14 economic predictors from Goyal and Welch (2008), output gap from Cooper and Priestley (2009), and short interest from Rapach et al. (2016). The last row reports the results using PLS with the 16 economic variables. Returns are in annualized percentages and the independent variable is standardized to zero mean and unit variance. Adjusted R^2 is in percentage; and t-stat is computed with the Newey and West (1987) standard errors. ***, **, and * indicate significant at 1%, 5%, and 10% respectively.

	Panel A: Univariate		Panel B: Bivariate			
Economic Predictor	$\gamma(\%)$	$R^2(\%)$	$\beta(\%)$	$\gamma(\%)$	$R^2(\%)$	Period
Dividend-price ratio (DP)	1.39	0.00	3.34 ***	0.81	0.28	187101-201910
Dividend yield (DY)	2.03	0.07	3.23 **	1.46	0.32	187102-201910
Earnings-price ratio (EP)	2.46	0.13	3.06 **	1.71	0.34	187101-201910
Dividend payout ratio (DE)	-1.00	-0.03	3.44 ***	-0.82	0.28	187101-201910
Stock variance (SVAR)	-0.08	-0.06	3.48 ***	-0.13	0.26	187101-201910
Book-to-market Ratio (BM)	5.19	0.58	2.56	4.60	0.64	192103-201910
Net equity expansion (NTIS)	-4.11	0.31	3.48 **	-3.76	0.51	192612-201910
Treasury bill rate (TBL)	-3.68 **	0.36	2.64 **	-2.92 *	0.50	187101-201910
Long term bond yield (LTY)	-2.82	0.11	3.05 *	-1.75	0.23	191901-201910
Long term bond return (LTR)	3.36 *	0.18	3.99 **	3.51 *	0.47	192601-201910
Term spread (TMS)	-2.70	0.10	3.10 *	-1.63	0.22	191901-201910
Default yield spread (DFY)	2.87	0.12	3.64 **	2.85	0.36	191901-201910
Default return spread (DFR)	2.30	0.04	3.84 **	2.27	0.30	192601-201910
Inflation (INFL)	-3.32	0.20	3.42 **	-4.07 *	0.41	191302-201910
Output Gap (OG)	-3.39	0.20	3.77 **	-3.59	0.47	191902-201910
Short Interest (SI)	-5.70 **	0.94	3.99	-5.75 **	1.30	197301-201412
Economic PLS	4.40 *	0.48	3.92	4.41 *	0.82	197301-201412

Table D4
Predicting Market Returns after Controlling for Uncertainty and Sentiment Variables Topic Weights Constructed by Raw Counts of Seed Words

This table presents the correlation between the narrative PLS index with each of the uncertainty and sentiment variables in Panel A, the results of the following predictive regression

$$R^e_{t+1} = \alpha + \gamma z_t + \epsilon_{t+1}$$

in Panel B, and the following predictive regression

$$R^e_{t+1} = \alpha + \beta x_t + \gamma z_t + \epsilon_{t+1}$$

in Panel C, where R^e_{t+1} is the excess market return over the next month, x_t is the narrative PLS index constructed by raw counts of seed words, and z_t is one of the uncertainty variables—financial and macro uncertainty indexes from Jurado et al. (2015), economic policy uncertainty index from Baker et al. (2016), disagreement index from Huang et al. (2020), and implied volatility (VIX)—and sentiment variables—news sentiment, investor sentiment from Baker and Wurgler (2006), aligned sentiment from Huang et al. (2015), and manager sentiment from Jiang et al. (2019). The last row reports the results using PLS with all uncertainty and sentiment variables. Returns are in annualized percentages and the independent variable is standardized to zero mean and unit variance. Adjusted R^2 is in percentage and t-stat is computed with the Newey and West (1987) standard errors. ***, **, and * indicate significant at 1%, 5%, and 10% respectively.

	Panel A: Correlations	Panel B: Univariate		Panel C: Bivariate			
Economic Predictor	Corr. with PLS (%)	γ(%)	R^2(%)	β(%)	γ(%)	R^2(%)	Period
Financial uncertainty	7.46 **	-5.75 **	1.15	4.65 **	-6.10 **	1.85	196007-201910
Macro uncertainty	5.42	-4.30	0.58	4.44 **	-4.54	1.20	196007-201910
Economic policy uncertainty	46.45 ***	4.03	0.38	2.63	2.81	0.35	198501-201910
Implied volatility (VIX)	17.94 ***	0.40	-0.27	5.27 **	-0.54	0.56	199001-201910
Disagreement	-11.95 ***	-8.43 ***	2.43	2.52	-8.12 ***	2.49	196912-201812
News sentiment	4.07 *	-0.52	-0.05	3.51 ***	-0.66	0.27	185701-201910
Investor sentiment (BW)	-4.63	-2.50	0.08	3.49 *	-2.34	0.38	196507-201812
Investor sentiment (PLS)	-10.65 ***	-7.32 ***	1.86	2.85	-7.02 ***	2.01	196507-201812
Manager sentiment	-38.18 ***	-9.06 ***	3.32	2.29	-8.19 *	2.99	200301-201712
Shiller's one-year confidence index	-37.15 ***	-4.77	0.48	7.70 **	-1.91	2.14	200107-201910
Shiller's crash confidence index	-40.30 ***	-2.07	-0.28	9.04 ***	1.57	2.10	200107-201910
Uncertainty PLS	43.70 ***	2.38	-0.29	5.41	0.02	0.26	200301-201712

Table D5
Out-Of-Sample R^2
Topic Weights Constructed by Raw Counts of Seed Words

This table reports the out-of-sample R^2_{OS} statistic (Campbell and Thompson, 2008) in predicting the monthly excess market return using the economic narratives constructed by raw counts of seed words. Panel A and B reports results using OLS and PLS, respectively. All of the out-of-sample forecasts are estimated recursively using data available in the expanding estimation window. All numbers are in percentages. ***, **, and * indicate 1%, 5%, and 10% significance of the Clark and West (2007) MSFE-adj statistic. The evaluation period begins in January 1891 and the whole sample is from January 1871 to October 2019.

	1891-2019	1891-1949	1950-2019	2000-2019
Panel A: OLS				
Dividend-price ratio (DP)	-0.39	-0.48	-0.25	0.05
Dividend yield (DY)	-0.34	-0.15	-0.64	0.04
Earnings-price ratio (EP)	-0.05	0.07	-0.26	-0.35
Dividend payout ratio (DE)	-0.65	-0.84	-0.33	-1.06
Stock variance (SVAR)	-1.67	-2.19	-0.79	-0.86
Treasury bill rate (TBL)	0.07 **	-0.04	0.26 **	0.45
Panic	0.08 ***	-0.08 **	0.36 **	0.75 *
Confidence	-0.18	-0.19	-0.16	-0.75
Saving	-0.15	-0.17	-0.11	-0.02
Consumption	-0.15	0.06	-0.52	0.11
Money	-0.16	-0.25	-0.01	-0.02
Tech	-0.40	-0.50	-0.21	-0.66
Real_estate	-0.19	-0.20	-0.18	-0.88
Stock	-0.09	-0.18	0.06	0.43
Boycott	-0.10	-0.18	0.02	0.30 ***
Wage	-0.15	-0.13	-0.20	0.19
Panel B: PLS				
Economic	-0.62	-0.76	-0.38	-0.55
Narrative	-0.16 **	-0.46 *	0.34 **	0.71 **