War Discourse and Disaster Premia: 160 Years of Evidence from Stock and Bond Markets*

David Hirshleifer† Dat Mai‡ Kuntara Pukthuanthong§

May 2023

Abstract

Using a semi-supervised topic model on 7,000,000 *New York Times* articles spanning 160 years, we test whether topics of media discourse predict future stock and bond market returns to test rational and behavioral hypotheses about market valuation of disaster risk. Focusing on media discourse addresses the challenge of sample size even when major disasters are rare. Our methodology avoids look-ahead bias and addresses semantic shifts. War discourse positively predicts market returns, with an out-of-sample R^2 of 1.35%, and negatively predicts returns on short-term government and investment-grade corporate bonds. The predictive power of war discourse increases in more recent time periods.

*We thank Fred Bereskin, Vineer Bhansali, Murray Frank (discussant), John Howe, Tim Loughran (discussant), Michael O'Doherty, Stephen Owen (discussant), Seth Pruitt (discussant), Richard Roll, Jeffrey Stark (discussant), Yinan Su (discussant), Russ Wermers (discussant), and the seminar and conference participants at the Bank of Portugal, the University of Cologne, the University of Missouri-Columbia, the University of Missouri-St. Louis, Missouri State University, the Doctoral Tutorial at the European Finance Association (2021), the Ph.D. 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), the American Finance Association PhD Student Poster Session (2022), the Midwest Finance Association Meeting (2022), the Young Scholars Finance Consortium (2022), the Boulder Summer Conference on Consumer Financial Decision Making (2022), and LongTail Alpha LLC for helpful comments. We acknowledge computing support from the Research Computing Support Services (RCSS) at the University of Missouri-Columbia.

†Marshall School of Business, University of Southern California, Email: hirshlei@marshall.usc.edu

‡Robert J. Trulaske, Sr. College of Business, University of Missouri-Columbia, Email: datmai@mail.missouri.edu

§Robert J. Trulaske, Sr. College of Business, University of Missouri-Columbia, Email: pukthuanthongk@missouri.edu

1 Introduction

We study here the asset pricing implications of the prospect of disasters, such as wars, pandemics, and political crises. Rare disaster risks are one of the leading possible rational explanations for major asset pricing puzzles (Barro 2006, 2009). A basic implication of rational disaster models is that high disaster risk will receive a risk premium, and therefore will predict high future stock market returns. Two behavioral hypotheses offer a similar implication. The first is attentional and belief-based: that investors overestimate the probability of rare disasters owing to the high salience of extreme outcomes. This possibility is supported by evidence that people overestimate the probabilities of rare events (Fischhoff, Slovic, and Lichtenstein 1977; Snowberg and Wolfers 2010). The second is preference-based: that investors overweight low probabilities, as in the expected value function of cumulative prospect theory (Tversky and Kahneman 1992).

The rarity of major disasters is a well-known obstacle to empirically testing the relationship between their occurrence and future stock returns.[1] This limits the power to identify effects. In this paper, we circumvent the obstacle of small sample size by using data on investors' attention to rare disaster risks derived from news. This provides a much larger sample of changing perceptions of disaster probabilities over 160 years.

Shifts in media topic coverage over time can potentially capture investor assessments of future prospects, including the risk of rare disasters. For example, Shiller (2019) argues that economic narratives are subject to occasional outbreaks when they spread rapidly and widely through the population, influencing behavior and decisions. He argues that such shifts can be captured by media discussions.

The novelty of our approach to capturing perceptions of disaster risk is twofold. First, we are the first study to compare the effects of disaster- and non-disaster-related topics of discourse systematically for the pricing of the aggregate stock market. Second, we apply a novel approach, *seeded Latent Dirichlet Allocation* (Lu et al. 2011, henceforth sLDA), to extract topics of popular discourse over time. This method has several key advantages over existing empirical application of

[1] An international political crisis occurs on average once every 15 years, a full-scale war once every 74 years, fighting on home territory once every 119 years, and a pandemic once every 100 years.

language models to asset pricing.

On the first point, for non-disaster-related topics, we consider the 12 "narratives" (topics) discussed in Shiller (2017, 2019). To obtain interpretable findings, rather than using a purely statistical procedure to extract topics, we draw upon the topics of Shiller (2019). We make necessary adjustments to these topics to effectively implement sLDA. The disaster-related topics we consider are war and pandemics.[2]

We find that non-disaster-focused topics predict stock market returns in-sample and have limited out-of-sample predictive power.[3] The *Pandemics* topic has limited and inconsistent predictive power even in-sample. *War* has the most predictive power both in- and out-of-sample. Our study is the first to show that war as a discourse topic is more powerful than non-disaster-focused topics in predicting returns.

On the second point, topic modeling is a prevalent dimension-reduction strategy in the machine learning and natural language processing literature that compresses large amounts of text into a limited set of topics. The topic model we use, sLDA, is a recent extension of the canonical unsupervised LDA model (Blei, Ng, and Jordan 2003; Griffiths and Steyvers 2004).[4] Since we use seven million articles published in the *New York Times (NYT)* over 160 years, it is crucial to apply a method that can process a large body of materials with low cost, reasonable speed, and limited error and subjectivity.

Under traditional unsupervised LDA, the model arbitrarily gathers common phrases and themes based on word frequencies. In contrast, under the semisupervised model or sLDA, the creation of themes allows control over the content of themes to be extracted. sLDA fits our research goal of testing the consequences of disaster-focused and non-disaster-focused themes in media discussions. In this approach, we feed the model with the seed words associated with each topic and let the

[2]Both can have massive human and economic costs and highly uncertain outcomes, as exemplified by the Covid-19 pandemic. Oleg Itskhoki, the winner of the 2022 John Bates Clark Medal, suggests in an interview with Bloomberg on August 2, 2022, that existing wars at that time presented an even greater economic risk than Covid (https://www.bloomberg.com/news/articles/2022-08-02/clark-medal-winner-oleg-itskhoki-says-war-is-a-bigger-economic-risk-than-covid?leadSource=uverify%20wall).

[3]The *Real Estate Booms* topic also yields significant out-of-sample predictive results, but since 1950, its out-of-sample predictive power is weaker that that of the *War* topic. *War* also outperforms the *Real Estate Booms* topic in terms of creating value for real-time investors. For details, see Section 6.

[4]LDA is burgeoning in popularity in computer science and other social science fields. For surveys, see Steyvers and Griffiths (2007), Blei (2012), and Boyd-Graber et al. (2017).

algorithm choose the phrases that often appear with these seed words.[5,6]

Our semisupervised topic model performs two key tasks: (1) it classifies market attention throughout the 160 years history of *The New York Times* into several disaster-focused and non-disaster-focused themes, and (2) it traces the evolution of media attention to these themes. These provide new quantitative measures of the market attention to topics of public discourse. Specifically, the model estimates the fraction of an article's text devoted to each topic. Aggregating over articles, these proportions measure the amount of news coverage each topic receives. This makes it feasible to assess, for instance, the relationship between news and asset returns.

Central to sLDA is the identification words that co-occur with the seed words. Our topic weights quantify the market's interest in each topic based on the frequency of terms that co-occur with it. We gather seed words from well-known media and publications, such as *Nature* for disaster-focused topics and from Shiller's book *Narrative Economics* for other topics. As we study 2 disaster-focused topics *War* and *Pandemic* and 12 non-disaster-focused topics discussed in Shiller (2019), we have 14 seeded topics in total. We add one unseeded topic to gather everything not captured by the seeded ones.[7]

There are two key challenges for estimating the predictive power of topics of discourse. First is the need to avoid look-ahead bias; second is the need to address the effects of semantic changes over time. To avoid look-ahead bias, parameter estimates at date t must be based only on the data available before date t.

This problem is intertwined with the second problem, that the meanings of words and phrases

[5] Recent papers in natural language processing, such as Lu et al. (2011), Jagarlamudi, Daumé III, and Udupa (2012), Eshima, Imai, and Sasaki (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 interpretability of topics and avoids the need to label extracted topics ex-post to interpret them.

[6] A third group of topic models is fully supervised methods (see for example Mcauliffe and Blei (2007) and Ramage et al. (2009)). These supervised models extract topics predictive of document tags or labels so they need labeled documents to train with. These models are not suitable for us because we want to use topics from news to predict market returns not article titles.

[7] Gentzkow, Kelly, and Taddy (2019) discuss the potential arbitrariness of the number of topics selected for study. In principle, one can employ optimization to determine the number of topics. However, doing so using data for the entire sample period would introduce look-ahead bias, which we seek to avoid. Furthermore, extracting the number of topics each month is unsuitable, as the common topic weights could fluctuate depending on each month's total topic count. Consequently, the weight of topics would not accurately represent market attention; it would instead vary due to the total number of topics for a given month. See also Subsection 5.8.

evolve over time.[8] Such semantic shifts are extensive since we use 160 years of data. An empirical approach that pools the entire sample to identify word lists and estimate the model parameters would not address such semantic shifts.

To address this issue, our analysis regularly updates the word list that constitutes topic weights. Although the list of seed words remains unchanged, the model is re-estimated monthly to address these two issues using data from the past ten years of news articles (including the current month). Therefore, the words clustered inside the topics reflect those popularly used during that time and vary monthly depending on semantic shifts. This ability to reflect semantic shifts over time is a key advantage of sLDA. In contrast, monthly rolling estimation is impossible under unsupervised LDA because it is not designed to yield consistent thematic content across estimations. This makes it impossible to address these two challenges.

Our estimation process addresses semantic changes by recalculating the topic weight on a monthly rolling forward basis. The output is an article-level weight vector with elements representing the proportion of content (or attention) devoted to the corresponding topic. We then compute the economy-wide monthly time series of topic weights for each topic from the article-level topic weights and use this aggregate time series in our tests of return predictability.

We use news media text to quantify perceived rare disaster risk. In principle, macroeconomic variables can be good proxies for the state variables of conditional asset pricing (Cochrane 1996). Empirically, however, macroeconomic variables perform poorly, with variation that is too low to match asset returns. News presents an alternative proxy of state variables that can potentially address this issue.

Our tests of conditional asset pricing are premised on state variables being captured by what investors read in the news. News is available at a high frequency comparable to that of asset returns.

[8]The meanings of words evolve over time spans of a century or more. For instance, "inflation" once referred to an increase in the money supply, but since the early 20th century it has referred to a general increase in the prices of goods and services in an economy (Homer and Sylla 1996). Another example is "amortization," which once referred only to the reduction of debt over time (Dictionary 1993), but in the 20th century has come to also refer to the gradual reduction in the value of an asset (Financial Accounting Standards Board (FASB) started this definition in 1973). The definition of "budget" has evolved substantially over time. In the 18th century, it referred to a financial statement outlining a government's anticipated expenses and revenues for the coming year. By the 1850s, the term expanded to include nongovernmental contexts, eventually encompassing the financial accounts of families or individuals (https://www.merriam-webster.com/words-at-play/financial-word-origins).

While the current literature on rare disaster risks uses contemporaneous price data (Ferguson 2006; Le Bris 2012; Oosterlinck and Landon-Lane 2006) to understand stock returns and bond returns during wartime, we use news, which allows us to continuously and quantitatively track the market attention to, or probability of, disaster risks.[9]

Our main results are based on data from the *NYT*; we also verify with robustness tests based on the *Wall Street Journal (WSJ)*. The *NYT* and the *WSJ* are the two largest national media outlets, and thereby can reflect the attention and perception of a large general audience.

Using all articles from *NYT* over 160 years provides several benefits.[10] First, although there is evidence that news about rare disasters affects investors' expectations and the equity return premium (see, for instance, Rietz (1988), Barro (2006), and Julliard and Ghosh (2012)), the rarity of extreme disasters makes it valuable to have a large enough sample to draw clear inferences. Second, long time-series data provides a testing ground for verifying the robustness of effects and tests whether the results are consistent over time (Schwert 1990). The *NYT* has been published since the 1860s, offering a comprehensive historical sample from a continuously available source. Therefore, we use data from the inception of the media to ensure a more accurate representation of the entire spectrum of news topics and their potential impact on stock market returns.

Also, using a broader dataset may be more representative of the topics of actual concern to investors. To our knowledge, this is the first paper that analyzes news articles from *all* newspaper sections of *NYT* since the beginning of its inception. Our sample includes nearly seven million articles from the *NYT* and six hundred thousand from the *WSJ*, making it a relatively comprehensive and extensive dataset.

Among the topics extracted from the *NYT*, we find that *War* is the strongest market predictor. The predictive power of *War* for the equity risk premium increases over time. Over 160 years, a one-standard-deviation increase in *War* predicts a 3.80% increase in annualized excess returns in the next month, and the monthly in-sample R^2 is 0.39%. In comparison, over the past 20 years, the

[9] Le Bris (2012) uses an event study approach to infer the effects of disasters. Berkman, Jacobsen, and Lee (2011) use a count of crisis events to proxy for rare disaster risk.

[10] We use the first ten years of news data to train our sLDA model to obtain the first month's topic weights and continue rolling the window forward. Our market index is unavailable from the Global Financial Data until 1871; thus, our training period starts in 1861, ten years after the *NYT*'s inception in 1851.

respective numbers are 9.83% and 3.39%. *War* is significant for both subperiods (1871–1949 and 1950–2019). This is economically substantial in comparison with the average annualized monthly excess stock market return over the same period, 3.96%. As another benchmark, the average R^2 of the 40 well-known predictors is only 0.73% in-sample and -1.01% out-of-sample (see Goyal, Welch, and Zafirov (2021), who discuss the weak out-of-sample performance of most economic return predictors). The R^2 of *War* indicates that its predictive power is economically significant. According to the time-varying disaster model, expected market excess returns should increase with the probability of rare disasters. Our results thus provide support for this theory.

In addition to *War*, we construct a discourse topic index from all 14 topics via the two-step Partial Least Squares (PLS) approach of Kelly and Pruitt (2013, 2015). The PLS index heavily loads on *War* with a correlation of 82%, so we interpret the PLS index as a more robust version of *War* that inherits its predictive power. Indeed, the monthly predictive regression of market returns with the PLS index yields a slope of 6.07% and an R^2 of 1.07% over the whole sample and a slope of 12.17% and an R^2 of 5.42% over the past 20 years. For the subperiods (1871–1949, 1950–2019), the PLS index remains a significant predictor at least the 1% level. The predictive power of the PLS index loaded mostly on *War* over market returns is not subsumed by common macroeconomic, sentiment, and uncertainty variables introduced in the literature.

We conduct standard out-of-sample tests as in the return predictability literature to investigate whether discourse topics create value for real-time investors. With expanding window estimation, *War* outperforms all individual economic 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. The out-of-sample R^2 of *War* is strong, with strongest return predictability in the last twenty years.

Turning to asset allocation implications, we consider a mean-variance investor who forms a portfolio allocating between stocks and a riskfree asset using either the return predictive model or the historical mean return to choose portfolio weights. Using *War* alone or combining topics to guide our portfolio decisions allows us to achieve a higher Sharpe ratio than a simple buy-and-hold strategy. Using a risk aversion coefficient of three following Huang et al. (2015) and Huang, Li, and Wang (2020), we find the economic gains for the investor utilizing discourse topics in forming portfolios increase over time, consistent with the R^2_{OS} results.

We also find that the perception of rare disaster risks predicts bonds' excess returns. In the model of Gabaix (2012), higher disaster risk implies a higher premium on long-term bonds. Our evidence is consistent with this prediction. We find that *War* positively predicts excess returns on mid- to long-term high-yield corporate bonds. Due to flight to quality, investors demand higher premiums to hold these high-yield assets. In contrast, *War* negatively predicts excess returns on safer investment instruments such as short-term government bonds and investment-grade corporate bonds. Our bond prediction results hold both in- and out-of-sample.

Contributions. This paper contributes to several lines of research. First, it contributes to social finance and narrative economics in testing how topics of media discourse are related to stock market pricing.

More specifically, our paper contributes to a line of research on the relationship of news content to economic and financial outcomes. In a related study that also uses the sLDA approach, Hirshleifer, Mai, and Pukthuanthong (2023) find that a factor that is based on our *War* variable explains the cross section of stock returns across a wide range of testing assets, and that leading benchmark factors as well as other media-based uncertainty measures do not subsume its explanatory power. Our paper differs in focusing on the time series predictability of the aggregate market return.

Our focus on predicting the aggregate market return distinguishes our paper from most existing research on news content and financial outcomes.

Bybee, Kelly, and Su (2023) aggregate disaster- and non-disaster-focused topics into a set of factors, where *Recession* is the most important topic for explaining the cross section of expected stock returns. Although not the main focus of their paper, they also report that their topics have little power to predict the aggregate market return. In contrast, our focus is on aggregate market return predictability. We consider disaster- and non-disaster-focused topics, and identify that *War* is a powerful predictor of aggregate market returns.

Applying data from many US local newspapers over a century, van Binsbergen et al. (2022) construct a measure of economic sentiment and find that it predicts future economic fundamentals, such as GDP, consumption, and employment growth. Our paper differs in comparing the effects of disaster- and non-disaster-related topics of discourse using media discourse to predict future stock and bond returns rather than economic fundamentals.

Perhaps the most closely related papers to ours are Adämmer and Schüssler (2020), Manela and Moreira (2017), and Bybee et al. (2023), all of which examine how news media can be used to predict aggregate stock market returns. Adämmer and Schüssler (2020) employ a variation of the unsupervised LDA model to extract topics from news articles and use them to forecast market returns. Our approach differs from theirs in three key respects. First, they focus on economic news in the *NYT* and *Washington Post* from 1980 to 2018, whereas our study utilizes all *NYT* sections starting from more than a century earlier. As emphasized by Lundblad (2007), since stock returns are highly volatile, it is crucial to consider long time series data to reliably test for return predictability.

Second, the method of Adämmer and Schüssler (2020), as with LDA, generates outputs that are challenging to interpret. Using statistical criteria to optimize over number of topics, the authors choose 100 topics for their model and identify one of them (topic 20) as the most important. The authors conclude that this topic represents geopolitical risk, but it is not obvious on prior conceptual grounds that these words would be the top choices for identifying such risk. (The most important words for this topic are "east," "west" and "German.") In contrast, our approach identifies investor concern with rare disasters. and in particular war, as crucial for return prediction.

Third, since they use an unsupervised topic model, their approach does not address semantic changes over time.[11] In contrast, our use of a semisupervised LDA model allows us to address semantic changes via monthly estimation of the topic model.

The top-line predictability reported in their study is a remarkable out-of-sample R^2 of 6.52%. However, their initial training window for return prediction of three years is too short for a reliable out-of-sample R^2 estimate.[12] In addition to finding a strong new predictor of market returns, our

[11] They train their topic model using news data from 1980 to 1995 and apply the trained model to extract topic weights from news articles from 1996 to 2018. They argue that because their sample is short, language change is not a concern (their footnote 7).

[12] The initial window used in our study is 10 years. An out-of-sample R^2 is computed by comparing the return forecast of a given model against the forecast using the historical mean return, which cannot be reliably estimated with only three years of data. Consistent with this, their out of sample R^2 is highly sensitive to different start dates

study identifies investor concern with disaster risk, and in particular war, as a determinant of asset pricing.

Manela and Moreira (2017) study how news events affect foreseen volatility and equity risk premia. They construct news implied volatility (NVIX) from the front page of *WSJ* starting from 1890. Using the data provided on the authors' website, we find that over 1900-2016 a standard deviation increase in NVIX^2 is associated with an increase in the annualized market excess return by 0.22% over the next month. A one standard deviation increase in our *War* variable is associated with an annualized market excess return of 2.7%.[13]

Our paper differs in several ways. First, our key return predictor is perceived disaster risk, as proxied by textual media, rather than general uncertainty. Perceived disaster risk could potentially be left tail risk rather than variance as studied by Manela and Moreira (2017).

Second, based on their support vector regression model training procedure, their risk measure captures only terms that appeared in the last 20 years of their sample period. This potentially induces look-ahead bias. In contrast, we estimate our model on a monthly rolling basis using data from the preceding 10 years. This allows us to address semantic changes over the 160-year sample period while avoiding look-ahead bias.

Third, we obtain stronger and more robust return predictability. We find that *War* predicts stock market return from one month to three years while their predictor, NVIX^2, does not predict returns until six months ahead. *War* is a powerful predictor in both in- and out-of-sample; Manela and Moreira (2017) do not report out-of-sample R^2.[14]

In independent work, Bybee et al. (2023) use traditional unsupervised LDA on news content to fit contemporaneous financial and macroeconomic variables and to forecast macroeconomic variables and stock market returns. They use cross-validation to select 180 topics in their LDA model. Our approach differs in using a semisupervised approach to focus on just 15 topics (14 seeded plus one

of the evaluation sample. As reported in their Table AV, moving the start date of evaluation period from 1999 to 1998 or to 2001 reduces the out-of-sample R^2 to 2.3%, which is comparable to our results over the 2000-2019 sample period.

[13] We report these results in Table 6. In their paper, Manela and Moreira (2017) report prediction results over 1945-2019 using NVIX^2.

[14] NVIX's in-sample predictive power almost disappears during the last twenty years of their sample period. Also, when we control for NVIX or NVIX^2, *War* is still a substantial and significant return predictor over the entire sample.

unseeded topic) to test hypotheses about rare disaster risk. Our analysis covers a much longer sample period, as we use all sections of the *NYT* from 1861 to 2019; Bybee et al. (2023) focus on economic news in the *WSJ* from 1984 to 2017. Finally, and crucially, our analysis avoids look-ahead bias by estimating our model on a rolling forward basis.[15]

Our paper also contributes to the literature on rare disaster risks, which incorporates disaster probabilities and loss into the standard consumption-based model to explain the high equity premium (Barro 2006, 2009; Gabaix 2012; Wachter 2013). Compared to other studies, we present the most comprehensive test of the predictions made by the time-varying rare disaster model or of the behavioral hypothesis that investors overweight rare disasters.

This paper is also related to the recent literature on extracting measures of political risk from textual data in relation to firm-level hiring and investment (see, e.g., Baker, Bloom, and Davis (2016); Hassan et al. (2019); Caldara and Iacoviello (2022)). Several studies, such as Pástor and Veronesi (2013) and Brogaard and Detzel (2015), document that the economic policy uncertainty (EPU) index in Baker, Bloom, and Davis (2016) positively predicts the aggregate market return over long horizons. In contrast, our *War* and PLS topic index are stronger predictors of one-month returns. So our index captures a different aspect of textual discourse and risk. Our time-series prediction results remain strong after controlling for the geopolitical risk measure (GPR) developed in Caldara and Iacoviello (2022), and yields stronger predictability for stock and bond returns than the dictionary approach of Caldara and Iacoviello (2022) (see Internet Appendix D).

Finally, our paper contributes to the burgeoning literature on applications of modern natural language processing tools to business and economic research. A growing body of research utilizes advanced topic modeling tools to extract thematic content from texts.[16] Unlike most finance papers that use the traditional unsupervised LDA model, our semisupervised LDA model allows us to extract a predefined set of topics in the news, which enhances interpretability.

[15] Bybee et al. (2023) address look-ahead bias by performing a rolling estimation of LDA via online LDA (oLDA). However, their use of 180 topics in the online LDA scheme was based on optimization over the entire sample period, and their use of a recession index in their out of sample test was informed by knowledge that the Recession topic was effective in the entire sample.

[16] See, for example, Dyer, Lang, and Stice-Lawrence (2017), Hansen, McMahon, and Prat (2018), Larsen and Thorsrud (2019), Choudhury et al. (2019), Brown, Crowley, and Elliott (2020), and Bybee et al. (2023).

2 Method

In this section, we briefly discuss the setup of the sLDA model (Lu et al. 2011) and our implementation of it to extract news topics.

2.1 The Stochastic Topic Model

Stochastic topic models are based on the core idea that documents can be described as mixtures of topics, where each topic is associated with a probability distribution over words (Blei 2012; Steyvers and Griffiths 2007). In this approach, latent topic weights are extracted from news articles. To do so, we assume that each text document is generated by a simple stochastic process that starts with a document-specific distribution over topics (the document-topic distribution). Each word in the document is chosen first by picking a topic randomly from the document-topic distribution and then drawing a word from the topic-word distribution for that topic.

The document-topic distribution for each document and topic-word distribution for each topic (the same across documents) are unobserved parameters that are estimated from the observable word frequencies in the document collection. We use standard statistical techniques to estimate the generative process, inferring the topics responsible for generating a collection of documents (Steyvers and Griffiths 2007).

The most widely used topic model is latent Dirichlet allocation (LDA) as introduced by Blei, Ng, and Jordan (2003) and further developed by Griffiths and Steyvers (2004). Under LDA, a document is generated under the hierarchical process described above. Each word in the document is selected by first randomly selecting a topic from the document-topic distribution, and then for that topic, the word is selected from the topic-word distribution. Under LDA, the document-topic distribution (a vector of probabilities over the topics) for each document is selected from a prior Dirichlet distribution (see Appendix A for details of the LDA and sLDA methodologies). The topic-word distribution is global; it does not depend on the document. It is also assumed to be drawn from a prior Dirichlet distribution. Since the topic-word distribution is a set of probabilities for drawing each possible word, the distribution for the number of instances of each word in an entire document is multinomial with these probabilities, with N being the number of words in the document.

The unknown parameters of the multinomial distributions are estimated using the frequencies of different words in the documents in the sample.[17] Specifically, we use Gibbs sampling to simulate the posterior distribution of words and documents and estimate the two hidden model parameters, namely the document-topic distribution (τ_d) and the topic-word distribution (ω_k).[18]

In traditional unsupervised LDA, only the number of topics K is prespecified; the model clusters words into these topics based on word frequencies in a completely unsupervised manner. The model automatically extracts underlying topics. The LDA model is more likely to assign a word w to a topic k in a document d if w has been assigned to k across many different documents and k has been used multiple times in d (Steyvers and Griffiths 2007).

Since we are interested in uncovering the effects of specific and interpretable topics relating to rare disaster risk, we instead employ a recent extension of LDA called seeded LDA (sLDA) developed by Lu et al. (2011). sLDA allows users to regulate topic contents using domain knowledge by injecting seed words (prior knowledge) into the model. When a seed word is not present in a text collection, it does not enter the sLDA model and has no impact on the estimation process.

2.2 Seed Words

A key component of an sLDA model is the set of seed words representing the prior knowledge of each topic. As emphasized by Watanabe and Zhou (2020), 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. Table 1 lists the lemmatized seed words for each topic. (Lemmatization is the removal of word endings such as *s, es, ing, ed*.) Our seed words for *War* include *conflict, tension, terrorism, war* and seed words for *Pandemic* include *epidemic, pandemic*.[19] The seed words need

[17] An exception is that the two hyperparameters of the two prior Dirichlet distributions are taken from LDA topic modeling literature.

[18] Gibbs sampling is a sampling technique to simulate a high-dimensional distribution by sampling from lower-dimensional subsets of variables where each subset is conditioned on the value of all others. See Griffiths and Steyvers (2004) for details on the implementation of Gibbs sampling in LDA.

[19] In the setup of LDA, "tension(s)" tends to not be assigned to *War* in documents that talk little about war (such as articles about tension headaches), and to be assigned to *War* in documents that talk a lot about war (such as

to be general fundamental concepts that have reasonably stable meanings over very long periods. Our methodology allows for the fact that the meanings of other words (such as "nuclear") may evolve over time or may even be neologisms that do not exist early in the sample.[20]

The seed words for the non-disaster-focused topics are manually collected from Shiller (2019). These words are discussed extensively in Shiller (2019). We also add certain words that help define the themes of the topics. Importantly, to avoid any look-ahead bias, in selecting the seed words, we exclude any words that were only introduced recently, such as *bitcoin*, *machine learning*, or *great recession*. As shown in Table 1, we have reclassified the 9 topics from Shiller (2019) into 12 topics to facilitate our estimation. Specifically, as *Panic* and *Confidence* are opposing notions, we split them into two topics. Similarly, *Frugality versus Conspicuous Consumption* is split into *Frugality* and *Conspicuous Consumption*. We further divide *Real Estate Booms and Bursts* into two separate topics, namely *Real Estates Booms* and *Real Estates Crashes*.[21] In addition to *Stock Market Bubbles*, we add *Stock Market Crashes*. In contrast, because of their similarities, we combine *Labor Saving Machines* and *Automation and Artificial Intelligence* into one topic.

In addition to the 14 topics discussed above, we include one additional "garbage collector" to absorb everything else in the news unrelated to these topics.

2.3 Estimation

Figure 1 illustrates the rolling estimation scheme used in the paper. At the end of each month t, we run the sLDA model using all news data over the preceding 120 months (months $t - 119$ to t). We use ten years of news data in the monthly estimation to balance the amount of news data required to estimate the model and computational costs. On average, every ten years of historical

articles about international tensions).

Our results remain robust, and *War* retains similar predictive power, when we include additional seed words such as *battle, front line, army, navy, weapons, military, officer, munitions, bombs, guns* and *battalion*.

[20] During the early 20$^{\text{th}}$ century, the term "nuclear" was primarily employed within the realm of atomic structure and nuclear physics (see Rutherford (2012)). As the mid-20$^{\text{th}}$ century approached, the development and utilization of nuclear weapons during World War II led to an association between "nuclear" and the immense destructive force of such armaments (Rhodes (2012)). In the aftermath of World War II and throughout the Cold War era, "nuclear" was increasingly linked to the application of nuclear technology for energy production (Walker (2004)). Advancing into the late 20$^{\text{th}}$ and early 21$^{\text{st}}$ centuries, the scope of "nuclear" broadened to encompass the concept of nuclear families (Cherlin (2010)).

[21] We replace the term "bursts" with "crashes," as the phrase "real estate burst" is not common in popular usage, and the word "burst" might be taken to mean a burst of positive activity, which is not the intended meaning.

data consists of around 460,000 articles, which should be sufficient to extract the topic weights at the time of estimation reliably. Notably, within topic models, rolling estimation is viable only under the sLDA model because the seed words that guide this approach provide consistency of thematic content over time.

We use Gibbs sampling to estimate the parameters of the model. We draw 200 drawings from the posterior distribution of z_{dv}, the realized topic for word location v in document d in the sLDA model, where we are conditioning on observed word frequencies.[22] In each drawing, we condition on the estimated values of the parameters of the model derived from previous drawing (where in the first draw, the initial estimate comes from a random number generator). In the last draw, we estimate our final value of the document-topic weights τ_d; that is, we estimate one 14×1 vector $\tau_d = [\tau_d^1, \tau_d^2, \ldots, \tau_d^{14}]$ for each news article, d, in the estimation window.

We then provide estimates of model parameters that condition on month t within the dataset. We compute the global monthly weights of each topic k ($k = 1, 2, \ldots, 14$) as the average weight of each topic across all articles in month t, weighted by the length $L(d)$ of each article:

$$\tau_t^k = \frac{\sum_{d=1}^{n_t} \tau_d^k L(d)}{\sum_{d=1}^{n_t} L(d)}, \tag{1}$$

where τ_t^k is the weight of topic k in month t, n_t is the total number of news articles in month t, and $L(d)$ is the total number of unigrams (one-word terms), bigrams (two-word terms), and trigrams (three-word terms) in article d.[23] (Equal weighting of topic weights across articles yields similar results.)

Although ten years of news articles are used to estimate the model each month, 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 14 topics. This time series is used for our economic and financial forecasting applications.

[22] In addition to the number of topics and articles, the number of samples drawn from the posterior distribution is a computational cost consideration in any topic model.

[23] An n-gram is a sequence of n words. For instance, "San Diego" is a bigram, and "A study of topics is needed" is a 6-gram.

3 Data

We leverage the richness of full newspaper texts using articles since the beginning of the *NYT* inception. We still remove articles with limited content, such as those that contain mostly numbers, names, or lists. We then conduct standard text processing steps, as reported in detail in the Internet Appendix B and Table C.1. Our analysis is based on all articles from the *NYT* since 1861 and all *WSJ* articles since 1990.[24,25] (This data was provided to us by a private company.)

Then, for each month t, we create a document term matrix containing all articles over the preceding ten years through the current month. Each row of the matrix is an article, each column is a n-gram, 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. To streamline the presentation, we report the results for the *WSJ* in Internet Appendix E.

Panel A of Figure C.1 plots the time series of monthly article counts in our sample. After removing articles with limited content, since 1871, our *NYT* data has more than 6.8 million news articles with a monthly average of 3,800.[26] Before 1900, the *NYT* published about 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 industry-wide struggles related to declining ad revenues and subscriber bases beginning in the 2000s, the *NYT* started scaling down its publishing capacity to around 2,000 articles a month during the 2010s.[27] 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.

[24] The *NYT* has been used in other finance research (see, e.g., Garcia (2013) and Hillert and Ungeheuer (2019)). The *NYT* has received 130 Pulitzer Prizes, almost double that of its nearest competitor. See https://www.nytco.com/company/prizes-awards/. Articles from the first ten years of the *NYT* since its inception are used to estimate the first monthly topic weights. We start our *NYT* sample in 1871 as the S&P 500 data is available from that year.

[25] Bybee et al. (2023) focus on economic articles in the *WSJ* with a sample that starts in 1984. Our sample starts one hundred years earlier. Manela and Moreira (2017) include the articles from the *WSJ* since 1890 but focus only on the headline and title of articles on the front page, whereas we cover all articles. Baker, Bloom, and Davis (2016) and Caldara and Iacoviello (2022) study an extensive collection of newspapers but apply a dictionary approach. Caldara and Iacoviello (2022) count, each month, the number of articles discussing rising geopolitical risks. They do not consider the number of words in each article whereas our methodology does.

[26] Data are missing for September and October 1978 (due to strikes) and thus are excluded from Figure C.1.

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

Panel B of Figure C.1 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). (Internet Appendix B provides more details on the construction of n-grams.) While Bybee et al. (2023) consider only unigrams and bigrams, we extend the analysis to trigrams as a majority of the seed words have three words. Examples include *real estate boom, stock market bubble*, and *cost push inflation*. Over 1871–2019, articles have an average length of 493 n-grams. Articles tended to have around 500 n-grams until the 1920s. After that, the number hovered just above 400 n-grams until the 1960s. Since then, article length has increased, reaching about 600 n-grams during the 2010s.

The second set of data concerns stock market outcomes. We obtain the total S&P 500 index from Global Financial Data (GFD) with monthly data from January 1871.[28] Monthly riskfree rates are downloaded from Professor Kenneth French's website. For monthly riskfree rates before 1927, we use the series from Goyal and Welch (2008).

4 Discourse Topics

We examine the contents of news topics in Subsection 4.1. In Subsection 4.2, we discuss the summary statistics of our topics and in Subsection 4.3, we discuss their time series.

4.1 Contents of News Topics

In a semisupervised topic model such as sLDA, the favored approach in the literature to evaluate the choice of seed words is to investigate the most common terms within a topic post-estimation to determine whether the topics feature the desired contents (Lu et al. 2011; Watanabe and Zhou 2020). Hence, to investigate the contents of the 14 extracted topics, during every monthly estimation of the sLDA model, we retain the 30 most common n-grams per topic, that is, those having the

[28]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."

highest probabilities in that topic. Then the most important words for each topic are identified as those that have the highest frequency over time.

To visualize each topic, we create word clouds using the top words from each topic; the higher the frequency of a word in the topic, the larger the word size. We report the word clouds of six main topics (based on their weights in the PLS index discussed next) in Figure 2, and the remaining topics in Figure C.2 in the Internet Appendix.

As indicated by Figure 2, the sLDA model seems to perform well at extracting these topics from the *NYT* articles. For example, the most common terms for *War* extracted by the model are *conflict, war, government, tension* and for *Panic* are *panic, fear, crisis, depression, recession, hard_time*, all of which strongly overlap with the seed words. Although both *War* and *Panic* feature stress and anxiety, they capture distinct themes, and their correlation is -17% as reported in Table C.2. The top words for *Monetary* are *money, gold, silver, inflation, bank*; for *Real Estate Booms* are *bubble, boom, speculation, price increase*; and for *Boycott* are *boycott, outrage, strike, moral, anger, community, protest.* Except for *Pandemic*, the thematic contents of these extracted topics are consistent with the predefined list of seed words.

4.2 Summary Statistics

We report the summary statistics for the 14 topic weights in Table 2. *War* receives the most attention on average, with a mean time-series weight of 9.7%. About 10% of the monthly *NYT* articles use one of the *War* words at least once. Table 2 shows that *War* is also the most volatile topic with a standard deviation of 3.7%, followed by *Stock Market Crash* at 3.1%.

For predicting stock market returns, we create a composite topic 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 (Huang, Li, and Wang 2020; Huang et al. 2015; Kelly and Pruitt 2013, 2015). As a first step, the time series of each topic weight is regressed on the time series of next-month 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. The construction of the PLS index for in-sample analyses uses the total sample from 1871 to 2019 as in the approach of Huang et al. (2015) and Huang, Li, and Wang (2020). For the out-of-sample analysis, we recursively reconstruct the PLS index every month using only the information available up to that month.

The second to last column of Table 2 reports the PLS loadings (the slope in the time-series regressions) for all topics. In this methodology, only the relative PLS weights of the components are meaningful. *War* receives the highest weight, and its positive loading indicates that *War* is a positive predictor of market returns. Other essential topics in the PLS index include *Real Estate Booms*, *Pandemic*, and *Panic*. Surprisingly, the topics receiving the smallest weights are *Stock Market Bubbles* and *Stock Market Crashes*. These facts are potentially useful for future theorizing about economic narratives and the stock market.

The last column of Table 2 reports the correlations between the 14 topics and the PLS index. As expected, the PLS index is highly correlated with *War* with a correlation coefficient of 82%.

4.3 Time Series of Discourse Topics

Next, we examine fluctuations in topic weights over time. We plot the time series of each demeaned topic weight against excess market returns from January 1871 to October 2019. The results for the six main topics and the PLS index are displayed in Figure 3, and the remaining topics are shown in Figure C.3 in the Internet Appendix. As can be seen from the graphs, except for *War*, the topics do not display any clear patterns. Thus, we focus our discussion on the time series of *War*.

Figure 3 describes the time series of *War*. *War* spiked in the 1870s, the Reconstruction period after the American Civil War. It also surged during the 1890s, a period that featured the Spanish-American War in 1898 and the Philippine-American War of 1899–1902. *War* rose to its highest since the start of the sample during World War I from 1917 to 1918. It remained low during the 1920s and 1930s before surging again during World War II. *War* reached its all-time high in 1963 due to major developments of the Vietnam War.

As shown later in this paper, the predictive power of *War* for the stock market has been stronger in recent decades. In Figure 4, we focus on the time series of *War* over the last 30 years of the sample. We track the ten articles with the most significant contributions to the ten highest monthly scores of *War* hikes since 1990.[29] Over the last 30 years, *War* spiked in the early 1990s during the Gulf War, and surged again at the end of 2001 after the 9/11 terrorist attack. In recent years *War* has remained high, especially from 2014 to 2018. During this time, the important articles reflect the climate of the period: stories are full of international tensions, notably including the Russian annexation of Crimea and the nuclear weapons threat from North Korea.

5 Discourse Topics as Stock Market Predictors

We next address the primary research question of the paper: do rare disasters and non-disaster-focused discourse topics predict the U.S. stock market return? In Subsection 5.1,we consider one-month return prediction. In Subsection 5.2 we consider long-horizon prediction. In the later subsections, we control for standard return predictors from past literature and examine whether discourse topics have incremental predictive power.

5.1 Predicting One-Month-Ahead Returns

To investigate the return predictive power of discourse topics, 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 topics or the topic PLS indexes standardized to zero mean and unit variance, and β, the coefficient of interest, measures return predictability. The reported t-statistics are computed with Newey and West (1987) standard errors.

Table 3 reports the results. Over the whole 1871-2019 sample, among the 14 topics, *War* is the strongest positive predictor, with the coefficient statistically significant at the 1% level. Economically, a one-standard-deviation increase in *War* is associated with a 3.8% increase in the

[29] Each month, the most influential article is the article with the highest product of article-level topic weight and article length, i.e., the numerator in Equation (1). Equal weighting, ignoring the article length, can help one identify slightly different influential articles. Still, these other articles are generally thematically similar to the most influential articles reported here.

annualized excess return in the next month.

In addition to the full sample analysis, we run predictive regressions over three subperiods: 1871-1949, 1950-2019 and 2000-2019. This addresses possible concerns about text quality in the earlier part of the sample. Furthermore, it is interesting to examine whether financial market behavior is different during the latest two decades, with the rise of internet usage and new communication technologies. The results during this period may be the most relevant for the future, as emphasized in Goyal and Welch (2008).

The positive association between *War* and future market returns remains in both subperiods with significance at the 5% level. Furthermore, *War* yields an impressive forecasting power over the past two decades with a coefficient of 9.8%, significant at the 1% level, and an in-sample R^2 of 3.4%.

Among the remaining economic discourse topics, *Pandemic* and *Real Estate Boom* are negative return predictors over the whole sample, both significant at the 5% level. In contrast, *Panic* is a positive predictor of market returns, significant at the 10% level. Among these non-war topics, only *Real Estate Boom* yields meaningful predictions across all subsamples.

The last portions of Table 3 report return prediction results using the PLS method. The PLS index constructed from all 14 topics predicts returns more strongly than *War* alone. Over the total sample, a one-standard-deviation increase in the PLS index is associated with a 6.1% increase in the annualized return in the next month, with an in-sample R^2 above 1%. Moving from earlier to later subsamples, the PLS index displays increasingly strong predictive results, significant at the 1% level even in the early subsamples. This suggests that the combined information in all topics has predictive power for long time-series data.

We also examine the predictive power of only the topics discussed by Shiller (2019). To do so, we construct the "Shiller PLS" index by excluding *War* and *Pandemic* and report the prediction results using this index in the last row of Table 3. Accordingly, the Shiller PLS index displays similar prediction patterns as the composite PLS index, albeit with smaller magnitudes. For example, over the whole sample, the Shiller PLS has a prediction coefficient of 4.8% and an R^2 of 0.6% compared to 6.1% and 1.1%, respectively, of the composite PLS index.

Following Golez and Koudijs (2018), we compute the cumulative in-sample R^2 in predicting the next month return, reported in Panel A of Figure 5. An upward trend indicates a predictor performs well during the sample period. Both *War* and the composite PLS index experience poor performances during 1910–1930 but strongly recover after that. Again, both cumulative R^2 suffer from a slight decline for a short period before 2000.

Overall, Table 3 indicates that *War* and the PLS index are strong market predictors, and their forecasting power increases in more recent periods. The predictive power of *War* and the PLS index is most pronounced from 2000–2019. We conjecture that the digitization of news and the technology that accelerates the diffusion of information drive this result. This result is consistent with that of Obaid and Pukthuanthong (2022), who find strong market predictive power in the sentiment of photos and text starting in 2010s.

5.2 Predicting Long-Horizon Returns

We have found that *War* and the PLS index predict market returns at a one-month horizon. We now examine the long-horizon predictive power of *War* and the PLS index by running the predictive regression:

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

where $R^e_{t+1\to t+h}$ is the annualized excess market return over the following h months, x_t is either *War* or the PLS indexes, and β, the coefficient of interest, measures the strength of predictability. In addition to the composite PLS index constructed using the whole sample reported in Table 4, to avoid look-ahead bias we also use an "expanding PLS" index computed every month using only the data available up to that month. To account for potential autocorrelations of the residuals in the long-horizon predictive regressions, we compute the Newey and West (1987) standard errors with corresponding h lags.

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 these 159 years, *War* and the two PLS indexes significantly predict market returns up to 36 months ahead. The PLS using the whole sample is built to optimize in-sample predictive power. As expected, the expanding PLS index yields smaller coefficient estimates and in-sample R^2.

In the subsample analysis, the predictive power of *War* is relatively weak during the first half of the sample period (significant at the 5% level for the one-month horizon). Still, it is significant at the 5% level from one- to six-month horizons during the second subperiod (1950 to 2019). The predictive power of the whole-sample PLS index is significant at the 1% level one month ahead and at the 5% level for other horizons (except 24 months) during the first half of our sample period. The results become stronger during the second half.

The strongest effects are obtained starting from the year 2000. *War* yields impressive predictive power over the last 20 years of the sample; its in-sample adjusted R^2 ranges from 3.4% (1 month) to 18% (36 months). During this period, the whole-sample PLS index yields strong results (significant at the 1% level) across all forecasting periods. As for economic magnitudes, a one-standard deviation increase in *War* is associated with an annualized increase of 9.8% in next month return over the 2000-2019 period. The corresponding number for the PLS index is 12.2%. The mean S&P500 annualized return during the same period is 3.96% suggesting the predictive power of *War* and the PLS index is economically substantial.

5.3 Predicting One-Month-Ahead Returns: *War* versus Economic and Topic Predictors

The previous two subsections show that *War* is a strong predictor of stock market returns. We next investigate whether *War* has predictive power beyond standard economic predictors and the remaining 13 topics studied in this paper. For economic predictors, we include the dividend-price ratio (DP), earnings-price ratio (EP), dividend payout ratio (DE), stock variance (SVAR), and T-bill rate (TBL) from Goyal and Welch (2008). We include these variables since they are available for our full sample period of 1871 to 2019. We run the following bivariate regression:

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

where z_t is either each of the economic or remaining topic predictors. All independent variables are standardized to zero mean and unit variance and t-statistics are computed with the Newey and West (1987) standard error.

Panel A of Table 5 reports the results for the economic predictors. In all bivariate regressions between *War* and each economic predictor, *War* remains significant at the 1% level with economic magnitude larger than those of the economic predictors. In the final column of Panel A, when we run a kitchen sink regression that includes *War* and all economic predictors,[30] *War* is still significant at the 5% level.[31]

In Panel B of Table 5, when *War* is tested against the remaining topic predictors, its statistical and economic significances remain intact in either bivariate or kitchen sink regressions.

Overall, we find that investors' perception of war risks as captured by our *War* index is a robust predictor of stock market returns and outperforms other common economic variables and non-disaster-related discourse topics in predicting the next one-month market returns.

5.4 Predicting One-Month-Ahead Returns: *War* versus Other Media-Based Uncertainty Indexes

In the previous subsection, we show that *War* as a measure of disaster risks has stronger predictive power than common economic predictors and non-disaster-focused discourse topics. However, the literature has introduced other news-based proxies for disaster risks, notably including the news implied volatility (NVIX) from Manela and Moreira (2017) and the geopolitical risks (GPR) from Caldara and Iacoviello (2022).[32] We now investigate whether our *War* contains incremental predictive power over these two measures.

In Panel A of Table 6, we use *War*, NVIX^2, and GPR to predict the excess market return one month ahead as in Equation (4). We first run univariate regressions and then compare *War* and GPR against NVIX^2. We do not directly compare *War* with GPR because the two variables are

[30] We exclude DE to avoid perfect collinearity because it is a linear combination of DP and EP.

[31] In Table C.3, we document that the predictive power of *War* remains intact when we control for market returns, conditional skewness, and conditional volatility.

[32] We thank the authors of these papers for making their data available.

highly correlated (correlation of 60%) over the sample period January 1900 to March 2016.[33] In contrast, *War* and NVIX^2 have a -5% correlation.

Over this period, both *War* and GPR are positive return predictors, significant at the 5% and 10% levels, respectively, where *War* yields a larger economic magnitude. In contrast, NVIX^2 does not predict the market at the one-month horizon, consistent with the results in Manela and Moreira (2017). We observe the same results over the sub-sample 1950-2016. Over the most recent sample, 2000-2016, *War* dominates in terms of both economic and statistical significance.[34]

Overall, *War* produces stronger short-term predictive power for market returns than the other two media-based disaster risk measures, especially after 2000.

5.5 Predicting One-Month-Ahead Returns: *War* versus Crisis Event Counts

In a previous study, Berkman, Jacobsen, and Lee (2011) measureinvestor perceptions of disaster risks by counting the number of crisis events each month. We now run a horse-race test of the effectiveness of this measure, which is based on actual crisis events, versus our media-based measure, which is based on textual discourse, in predicting returns.[35] We include the aggregate crisis index as this is the main variable studied in Berkman, Jacobsen, and Lee (2011) and the war count index as this is mostly related to our *War*. In Panel B of Table 6, we show that over the 100-year period 1918-2018, both news-based (*War*) and event-based (*CWar*) war indexes predict next-month market returns, both significant at the 5% level. However, over the two subsamples 1950-2018 and 2000-2018, *War* dominates the event count indexes.[36]

We conclude that investors' perception of rare disaster risks as extracted from news media is an important predictor of stock market return, even after controlling for event-count variables.

[33] NVIX is only available until March 2016. Following their paper, we use NVIX^2 in our analyses. Using NVIX yields almost the same results.

[34] In unreported results, when we put three predictors together in the same regression, *War* drives out the significance of GPR during the period 2000-2016.

[35] The data is updated to 2018 and available at https://sites.duke.edu/icbdata/.

[36] In Appendix C, we explore a larger set of real crisis events obtained from Global Financial Data. We create dummy variables to capture the occurrences of the following events: recessions, bank failures, wars, natural disasters, epidemics, and any of them. As reported in Table C.5, *War* retains its predictive power after controlling for these events.

5.6 Predicting One-Month-Ahead Returns: Controlling for Economic Variables

In the previous subsections, we show that *War* outperforms 5 economic variables, 13 topics, and numerous crisis event count as indexes in predicting next month stock returns. In this subsection, we extend the list of economic variables and consider the following bivariate predictive regression:

$$R^e_{t+1} = \alpha + \beta x_t + \gamma z_t + \epsilon_{t+1}, \tag{5}$$

where z_t is one of the economic predictors. Following Huang, Li, and Wang (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, Ringgenberg, and Zhou (2016).

The "Univariate" column in Panel A of Table 7 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. At the same time, the long-term bond return is the only significant positive predictor, marginally significant at the 10% level. The last row reports the prediction results with a PLS index constructed with 16 economic variables (hereafter, the economic PLS index). *War is* significant at the 10% level.

In the "Bivariate" column of Panel A, the topic PLS index is tested against each economic predictor in bivariate regressions. The PLS index is used instead of *War* as the former inherits the latter's features and is a stronger predictor. The PLS index remains significant at the 5% level or stronger against the 16 economic predictors. Finally, when tested against the economic PLS index, the topic PLS index remains significant at the 10% level and drives out the significance of the economic index. Overall, the results indicate that the discourse topics contain substantial information for predicting market returns beyond that in standard economic predictors.

5.7 Predicting One-Month-Ahead Returns: Controlling for Uncertainty and Sentiment Variables

The previous section documented that the topic PLS index contains valuable insights into market returns. We now test whether the topic PLS index has incremental ability to predict returns in comparison with other well-known uncertainty or sentiment variables. Recently, ample measures have been introduced into the literature, notably the financial and macro uncertainty indexes from Jurado, Ludvigson, and Ng (2015), the economic policy uncertainty index from Baker, Bloom, and Davis (2016), and the disagreement index from Huang, Li, and Wang (2020). Another commonly used measure of uncertainty is the Chicago Board Options Exchange's Volatility Index (VIX).

Another strand of the predictability literature studies sentiment measures. The most influential of these is the investor sentiment index of Baker and Wurgler (2006) (hereafter BW sentiment), which 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). Jiang et al. (2019) construct manager sentiment from corporate filings to show that manager sentiment has predictability beyond what is captured by investor sentiment. Moreover, Tetlock (2007) and Garcia (2013) find that sentiment extracted from news articles predicts daily market returns. To construct news sentiment from the *NYT*, we 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 also include the two U.S. stock market confidence indexes introduced by Shiller: the one-year confidence index and the crash confidence index.[37]

The "Correlations" column of Panel B of Table 7 reports the pairwise correlations between the topic index and each uncertainty and sentiment index. The topic PLS index has a significant 26% correlation with the economic policy uncertainty in Baker, Bloom, and Davis (2016) and a significant -19% correlation with the manager sentiment index in Jiang et al. (2019). Furthermore,

[37]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.

the topic PLS index is significantly negatively correlated with Shiller's one-year confidence index (correlation -30%).

The "Univariate" column of Panel B reports the univariate prediction for each uncertainty and sentiment variable. We find that the disagreement index, PLS sentiment index, and manager sentiment index show strong prediction results, consistent with previous papers. The financial uncertainty index by Jurado, Ludvigson, and Ng (2015) is a negative predictor, significant at the 5% level. The economic policy uncertainty index of Baker, Bloom, and Davis (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)).

In the "Bivariate" column of Panel B, we test the topic PLS index against the other sentiment and uncertainty variables. The PLS index remains significant in each multivariate predictive regression (at the 10% level or stronger). The last row of Table 7 reports the results with the PLS index constructed from all the uncertainty and sentiment variables, against which the PLS index remains significant at the 5% level.

The results in Table 7 indicate that the discourse topic index contains valuable information about market returns after controlling for the strong market predictors recently proposed in the literature.

5.8 Robustness Checks

We next consider the robustness of our results with respect to several features of the empirical design. Lu et al. (2011) suggest using a number of topics equal to the number of seeded topics plus one.[38] It would not be appropriate to optimize the number of topics using the entire sample period to maximize ex post predictive power, as this would expose the study to look-ahead bias. Another possibility might be to use statistical methods such as Bayes factors or cross-validation to select the optimal number of topics during each monthly estimation. However, under such an approach,

[38] The rationale for adding an unseeded topic is to allow the model to discover and consider an additional topic that may not be captured by the seeded topics provided. By using the number of seeded topics plus one, the model can better accommodate any unforeseen or unanticipated topics that are relevant to the data but were not part of the initial seeded topics. This approach helps strike a balance between incorporating prior knowledge through seeded topics and remaining flexible enough to account for any new information or patterns that may arise from the data during analysis.

the optimal numbers of topics can change from month to month. As a result, monthly changes in topic weights would not be attributable to the shifts in public attention to different topics that are central to our approach.

In this paper, since we are interested in studying 2 disaster-focused topics and 12 non-disaster topics from Shiller (2019), we only include 15 topics in sLDA (14 seeded plus 1 unseeded topic). Our exogenous specification of the number of topics is consistent with the view in Gentzkow, Kelly, and Taddy (2019) that in many applications of topic models, the goal is to provide an intuitive description of text rather than infer underlying "true" parameters. Our out-of-sample return predictability tests provide a validation that our specification of the number of topics generates meaningful results.

To evaluate the robustness of our findings with respect to the number of topics and the number of seed words, we focus on the 2000-2019 sample since this period shows the strongest return predictive power. For the number of topics, we experiment by increasing the number of unseeded topics to 50 while keeping the seeded topics unchanged and find that the predictive power for *War* is robust.

For the number of seed words, Lu et al. (2011) do not provide a guideline but mention in their footnote 6 that the number of seed words can be expanded depending on the size of the topic. We try varying numbers of seed words, duplicates in seed words within and across the topics, and the types of seed words such as unigrams, bigrams, or trigrams. We experiment with each of these choices subsequently while keeping other specifications of the model unchanged and find that the predictive power of *War* remains robust.[39]

As detailed in Internet Appendix B, we create n-grams within punctuation boundaries before removing stop words. We then remove bigrams containing stop words because these bigrams add no additional value to topic estimation beyond the unigrams contained therein. We also remove trigrams containing stop words unless the stop word is a preposition in the middle position. Our n-gram creation method ensures we only consider meaningful terms in the text data. However, we

[39]Specifically, we first vary the number of seed words while keeping other specifications unchanged. We then vary duplicates of seed words while keeping other specifications unchanged. Finally, we vary the types of seed words and while keeping everything else unchanged.

also experiment with removing stop words *before* creating n-grams and retaining all the resulting n-grams and find consistent results for *War* over 2000-2019.

When cleaning the texts, we remove articles containing mostly numbers, names, and lists (i.e., articles having limited content) by manually examining the patterns of their titles. As a robustness check, we keep all news articles in our data set and find qualitatively similar results for *War*.

We also examine the strategy of using a very large number of topics. In such a case, the weights of seeded topics can be approximated by the frequency of the seed words in the corpus. So, we investigate this case by constructing topic weights as the counts of seed words scaled by the article length and present the results in Internet Appendix D. Frequency-based topic weights still yield results consistent with the sLDA ones, but their out-of-sample performance is weaker.

6 Out-of-Sample Analysis

The predictability results in Section 5 are obtained by pooling within the 150-year sample. Such tests are subject to look-ahead bias, as with past studies that perform in-sample predictability tests or that use in-sample information to construct return predictors. To address this concern, we now perform an out-of-sample analysis, as is required to offer real-time economic value to investors (Goyal and Welch 2008). We conduct two standard out-of-sample tests to investigate whether discourse topics can help investors make better investment decisions: out-of-sample R^2 and certainty equivalent return (CER) gains.

6.1 Out-of-Sample R-Squared

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

$$R^2_{OS} = 1 - \frac{\sum_{t=p}^{T-1} \left(R^e_{t+1} - \hat{R}^e_{t+1} \right)^2}{\sum_{t=p}^{T-1} \left(R^e_{t+1} - \bar{R}^e_{t+1} \right)^2}, \tag{6}$$

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 1881 (10 years from the sample's start).

We benchmark the out-of-sample results of the 14 topics against the six 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 two approaches to estimate the function $f_t(x_t)$ recursively. First, we specify $f_t(x_t)$ to be a linear function of the 14 topics and 6 other predictors. Second, we specify $f_t(x_t)$ to be a function of all 14 topics or all 6 other predictors estimated via PLS as described in Section 4. Also, recall that our topic weights are extracted monthly using data over the past ten years, so there is no look-ahead bias in the out-of-sample analysis.

When a predictor outperforms the historical mean benchmark in forecasting future returns, it produces a lower mean squared forecast error (MSFE) than the historical mean. 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 8 reports the results from OLS regressions 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 14 topics, during 1881–2019, *War*, *Pandemic*, and *Real Estate Boom* yield a significant R^2_{OS} (0.17%, 0.08%, and 0.19%, respectively). Except for *Pandemic*, *War* and *Real Estate Boom* continue to deliver out-of-sample (henceforth, OOS) predictive power over the past 20 years with magnitudes much larger than the whole-sample results, at 1.35% and 1.13%, respectively. Consistent with the in-sample results in Section 5, *War* displays strong out-of-sample predictive power in recent periods.

Panel B of Table 8 combines the signals of individual predictors via PLS. Combining all six other predictors produces negative R^2's across all sample periods. Over the whole sample, combining all topics via PLS yields a negative R^2_{OS}. However, in the two most recent subsamples, the topic PLS method delivers strong predictive power, producing R^2_{OS}'s of 0.95% over 1950–2019 and 2.23% over

2000–2019, both significant at the 1% level. The predictive power of the topic PLS provides an economically substantial superior performance compared to the R^2_{OS} of 1.7% from Gómez-Cram (2022) using macroeconomic indicators over the last twenty years. In the last row of Panel B, we use only the 12 topics from Shiller (2019) in the PLS estimation, which yields negative R^2_{OS}'s in all samples.

Panel B of Figure 5 plots the cumulative out-of-sample R^2 for *War* and the PLS method using all 14 topics. An upward trend indicates good performance during that period. Consistent with Table 8, *War* and the PLS method do not perform well during the first half of the sample, especially from 1910 to 1930, in which both display a steep downward slope in the cumulative R^2_{OS}. From 1930 to 1990, both *War* and the PLS method feature steadily upwards trends, with the PLS method having a much steeper slope. However, both encounter a decline during the 1990s before having a turnaround during the last two decades of the sample.

Overall, we find that discourse topics outperform standard return predictors in out-of-sample prediction, especially during recent decades. These findings corroborate the in-sample results of earlier sections that the predictive power of discourse topics is stronger in recent periods.

Recent research has documented that sentiment variables have stronger predictive power during recessions and high sentiment periods when mispricing is likely to be prevalent owing to short-sale constraints (see, e.g., Garcia (2013), Huang et al. (2015), and Jiang et al. (2019)). We investigate whether our news topics follow the same pattern by decomposing the whole sample into expansions and recessions as well as high and low sentiment periods and compute the in- and out-of-sample R^2's during each subperiod. As reported in Table C.6, discourse topics better predict the market return during low sentiment periods. However, we find no evidence of different predictive powers across the business cycles. These results highlight that the predictive power of discourse topics operates via a different channel from that of sentiment.

6.2 Asset Allocation Implications

We next examine the economic value of news topics from an asset allocation perspective. Following Campbell and Thompson (2008), we compute the certainty equivalent return (CER) gain and

Sharpe ratio for a mean-variance investor who optimally allocates her portfolio between the stock market and the riskfree asset using out-of-sample return forecasts.

At the end of period t, the investor optimally allocates

$$w_{t+1} = \frac{1}{\gamma} \frac{\hat{R}^e_{t+1}}{\hat{\sigma}^2_{t+1}} \tag{7}$$

of the portfolio to equities during period $t+1$, where risk aversion coefficient γ is set to three following Huang, Li, and Wang (2020),[40] $\hat{R}^e_{t+1}$ 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 riskfree asset. The $t+1$ realized portfolio return is

$$R^p_{t+1} = w_{t+1} R^e_{t+1} + R^f_{t+1}, \tag{8}$$

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

The CER of the portfolio is

$$CER_p = \hat{\mu}_p - 0.5\gamma\hat{\sigma}^2_p, \tag{9}$$

where $\hat{\mu}_p$ and $\hat{\sigma}^2_p$ are the samples mean and variance, respectively, for the realized portfolio returns over the evaluation period. The CER gain is 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 annual management fee the investor is willing to pay to gain access to the predictive forecasts. In addition to the CER gain, we 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, Garlappi, and Uppal (2009).

Panel A of Table 9 reports the asset allocation results when individual predictors are used to

[40] To conserve space, the results with a risk aversion coefficient of five are not reported but are similar to the reported results.

make return forecasts. Treasury Bill produces the highest utility gains among all predictors during 1881–2019 at 1.31%, while *Money* comes in second at 1.03% and *War* comes third at 0.86%. While Treasury Bill continues to deliver utility gains over each subsample, over the past 20 years, EP (the earnings price ratio) and *War* produce better allocation performance at 3.88% and 2.01%, respectively.

Panel B of Table 9 shows that over 2000-2019, results via PLS yield a utility gain of 4.11%, the highest among all setups considered. Consistent with the OOS R^2 results, using all 14 topics yields superior performance to utilizing only narratives from Shiller (2019).

The right panel of Table 9 shows the results for the annualized Sharpe ratio. Among all the individual predictors in Panel A, Treasury Bill yields the best results for the whole sample, followed by *Money* and *War*. Combining all economic predictors or topics via PLS only delivers significant results from 2000-2019. As a benchmark, the last row reports the annualized monthly Sharpe ratio from buying and holding the S&P 500 index in the corresponding periods. Allocations using the combination of topics outperform the buy-and-hold strategy in general.

Overall, the allocation results suggest that using return forecasts from *War* or the combination of discourse topics via PLS offers real-time economic values to investors. The economic gains increase over time, consistent with the R^2_{OOS} results.

7 *War* and Predictability of Bond Returns

We have shown that *War* is a positive stock market predictor. This is consistent with the disaster risk model (Barro 2006, 2009), or with a behavioral model in which rare risks are overestimated or overweighted in investors' expected utility functions. Gabaix (2012) theoretically shows that disaster risks should also affect bond risk premia. Specifically, disaster probabilities should increase risk premia on risky bonds such as long-term high-yield corporate bonds and decrease risk premia on safe bonds such as short-term government and investment-grade corporate bonds. A behavioral setting in which investors overweight rare risks suggests a similar prediction. We now test these predictions.

Specifically, we run the following predictive regression:

$$R^e_{t+1\to t+h} = \alpha + \beta War_t + \epsilon_{t+1\to t+h}, \tag{10}$$

where $R^e_{t+1\to t+h}$ is the annualized excess returns over the next h months on Treasury bond indexes, investment grade, and high-yield corporate bond indexes, and War_t, as before, is *War* standardized to zero mean and unit variance. We consider the prediction horizons of 1, 3, 6, and 12 months ahead. The Treasury bond indexes are from Datastream and available from December 1988 to October 2019; the corporate bond indexes are from S&P and available from January 1993 to October 2019. To measure the statistical strength of β, we report the Newey and West (1987) standard error with h lags.

Panel A of Table 10 reports the results for Treasury bonds. We consider first the whole sample 1988-2019 over which the data for Treasury bond indexes are available. During this period, *War* does not have a discernible impact on government bonds' excess returns. We next consider the post-2000 period, in which *War* displays the most robust predictive power for stocks. During this sample, *War* negatively predicts excess returns over the next 1 up to 12 months ahead on short-term Treasury bonds having a maturity of up to 7 years. For longer-maturity Treasury bonds, *War* has no predictive power.

For investment grade corporate bonds in Panel B, over 1993-2019, *War* is a negative predictor of excess returns of bonds maturing in under one year, marginally significant at the 10% level. From 2000-2019, the impact is more substantial and powerful than in the longer sample period, for bonds with a maturity of 3 years or less.

For high-yield corporate bonds in Panel C, *War* positively predicts excess returns on bonds having 3 years or more until maturity over the whole sample 1993-2019. The most robust predictability is found in 5-7 year high yield bonds, significant at the 5% level for the next month's return. For the 2000-2019 sample, *War* is still a positive predictor of subsequent one-month excess returns on high-yield corporate bonds ranging from 3 to 10 years to maturity.

We also perform the out-of-sample predictions of bond excess returns using *War* and report the results in Table 11. To facilitate comparison among different types of bond indexes, we limit the sample to the range of January 1993 to October 2019. We employ an expanding estimation window using the first 10 years of the sample as the initial estimation window. We find, consistent with the in-sample results, that *War* has out-of-sample predictive power for returns of short-term government bonds (1-3 years to maturity) and investment-grade corporate bonds (0-1 years to maturity). *War* also yields significant R^2_{OS}'s for following one-month returns on high-yield corporate bonds having 3 to 10 years of maturity.

Overall, our bond prediction results are consistent with the predictions of the rare disaster risk model and with behavioral models in which rare risks are overweighted by investors. While the empirical disaster probability captured by *War* is a negative predictor of safe assets such as short-term government bonds and investment grade corporate bonds, it is associated with an increase in the return premia of risky investments such as stocks and mid- to long-term high-yield corporate bonds. The absolute coefficients on high-yield bonds are about seven to 15 times larger than those of investment-grade bonds with the same maturity.

8 Conclusion

We test the hypothesis that rare disaster risk is priced (or mispriced) by extracting market attention to rare disasters from news media. This helps overcome the challenge of scant data on realized disasters. It also has the advantage (from a behavioral perspective) of focusing on investor attention to and perceptions of disaster, which may differ from objective risks. We provide the most comprehensive analysis for empirically testing for pricing effects of disaster risk. In addition to two topics covering rare disaster risks (*War* and *Pandemic*), we also examine 12 non-disaster-focused narratives from Shiller (2019).

We employ an advanced natural language processing tool called sLDA to extract discourse topics from nearly seven million *New York Times* articles over the past 160 years. We create a list of topic-based seed words to input into the sLDA model to guide the topic extraction process. We employ a rolling estimation scheme to include only historical news data at every estimation time; thus, our measure avoids look-ahead bias and addresses changes in semantic usage over time.

Among the discourse topics considered, the most important is *War*, which encompasses various themes related to the danger of armed conflict. We find that *War* and an index constructed from all topics are strong positive predictors of the stock market return. We find the predictive power of *War* increases through the sample period and the predictive power of discourse topics holds at both the market and portfolio levels.[41] The predictive power of *War* remains even when extracted from a different media outlet, the *WSJ*.

That *War* positively predicts excess market returns is consistent with the literature on rare disaster risks, or with the behavioral hypothesis that investors overweight the prospect of rare disasters. Barro (2009) finds that the probability of rare disasters can explain the high equity premium. During times when the probability of a rare disaster is higher, the equity premium should be higher, which is consistent with our finding that *War* is associated with higher subsequent stock returns. Alternatively, if investors overweight rare risks (either owing to overestimation of probability owing to salience or the overweighting of low probability events in the cumulative prospect theory utility function), we again expect higher war media discourse to be associated with high future returns.

Our results also confirm the prediction of Gabaix (2012) (or of a behavioral setting where agents overweight rare disasters). We find that *War* increases the excess returns on mid- to long-term high-yield corporate bonds. In contrast, *War* negatively predicts excess returns on safer investment instruments such as short-term government bonds and investment-grade corporate bonds.

[41] See Table C.7 for portfolio prediction results.

References

Adämmer, P., & Schüssler, R. A. (2020). Forecasting the Equity Premium: Mind the News! *Review of Finance*, *24*(6), 1313–1355.

Baker, M., & Wurgler, J. (2006). Investor Sentiment and the Cross-Section of Stock Returns. *Journal of Finance*, *61*(4), 1645–1680.

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

Barro, R. J. (2006). Rare Disasters and Asset Markets in the Twentieth Century. *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.

Berkman, H., Jacobsen, B., & Lee, J. B. (2011). Time-Varying Rare Disaster Risk and Stock Returns. *Journal of Financial Economics*, *101*(2), 313–332.

Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, *55*(4), 77–84.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (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.

Boyd-Graber, J., Hu, Y., Mimno, D., & others others, o. (2017). Applications of Topic Models. *Foundations and Trends in Information Retrieval*, *11*(2-3), 143–296.

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

Brown, N. C., Crowley, R. M., & Elliott, W. B. (2020). What are You Saying? Using Topic to Detect Financial Misreporting. *Journal of Accounting Research*, *58*(1), 237–291.

Bybee, L., Kelly, B. T., & Su, Y. (2023). Narrative Asset Pricing: Interpretable Systematic Risk Factors from News Text. *Review of Financial Studies (forthcoming)*.

Bybee, L., Kelly, B. T., Manela, A., & Xiu, D. (2023). Business News and Business Cycles. *Working Paper*.

Caldara, D., & Iacoviello, M. (2022). Measuring geopolitical risk. *American Economic Review*, *112*(4), 1194–1225.

Campbell, J. Y., & Cochrane, J. H. (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., & Thompson, S. B. (2008). Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average? *Review of Financial Studies*, *21*(4), 1509–1531.

Cherlin, A. J. (2010). *The Marriage-go-round: The State of Marriage and the Family in America Today*. Vintage.

Choudhury, P., Wang, D., Carlson, N. A., & Khanna, T. (2019). Machine Learning Approaches to Facial and Text Analysis: Discovering CEO Oral Communication Styles. *Strategic Management Journal*, *40*(11), 1705–1732.

Clark, T. E., & West, K. D. (2007). Approximately Normal Tests for Equal Predictive Accuracy in Nested Models. *Journal of Econometrics*, *138*(1), 291–311.

Cochrane, J. H. (1996). A cross-sectional test of an investment-based asset pricing model. *Journal of Political Economy*, *104*(3), 572–621.

Cooper, I., & Priestley, R. (2009). Time-Varying Risk Premiums and the Output Gap. *Review of Financial Studies*, *22*(7), 2801–2833.

DeMiguel, V., Garlappi, L., & Uppal, R. (2009). Optimal versus Naive Diversification: How Inefficient is the 1/N Portfolio Strategy? *Review of Financial Studies*, *22*(5), 1915–1953.

Dictionary, O. E. (1993). Oxford English Dictionary. *Simpson, JA & Weiner, ESC.–1989*.

Dyer, T., Lang, M., & Stice-Lawrence, L. (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., & Bollerslev, T. (1986). Modelling the Persistence of Conditional Variances. *Econometric Reviews*, *5*(1), 1–50.

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

Ferguson, N. (2006). Political Risk and the International Bond Market between the 1848 Revolution and the Outbreak of the First World War. *Economic History Review*, *59*(1), 70–112.

Fischhoff, B., Slovic, P., & Lichtenstein, S. (1977). Knowing with Certainty: The Appropriateness of Extreme Confidence. *Journal of Experimental Psychology: Human Perception and Performance*, *3*(4), 552.

Gabaix, X. (2012). Variable Rare Disasters: An Exactly Solved Framework for Ten Puzzles in Macro-Finance. *Quarterly Journal of Economics*, *127*(2), 645–700.

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

Gentzkow, M., Kelly, B., & Taddy, M. (2019). Text as Data. *Journal of Economic Literature*, *57*(3), 535–74.

Golez, B., & Koudijs, P. (2018). Four Centuries of Return Predictability. *Journal of Financial Economics*, *127*(2), 248–263.

Gómez-Cram, R. (2022). Late to Recessions: Stocks and the Business Cycle. *The Journal of Finance*, *77*(2), 923–966.

Goyal, A., & Welch, I. (2008). A Comprehensive Look at the Empirical Performance of Equity Premium Prediction. *Review of Financial Studies*, *21*(4), 1455–1508.

Goyal, A., Welch, I., & Zafirov, A. (2021). A Comprehensive Look at the Empirical Performance of Equity Premium Prediction II. *SSRN*.

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

Hansen, S., McMahon, M., & Prat, A. (2018). Transparency and Deliberation within the FOMC: A Computational Linguistics Approach. *Quarterly Journal of Economics*, *133*(2), 801–870.

Hassan, T. A., Hollander, S., van Lent, L., & Tahoun, A. (2019). Firm-Level Political Risk: Measurement and Effects. *Quarterly Journal of Economics*, *134*(4), 2135–2202.

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

Hirshleifer, D., Mai, D., & Pukthuanthong, K. (2023). War Discourse and the Cross-Section of Expected Stock Returns. *Working Paper*.

Homer, S., & Sylla, R. E. (1996). *A History of Interest Rates*. Rutgers University Press.

Huang, D., Li, J., & Wang, L. (2020). Are Disagreements Agreeable? Evidence from Information Aggregation. *Journal of Financial Economics*, *141*(1), 83–101.

Huang, D., Jiang, F., Tu, J., & Zhou, G. (2015). Investor Sentiment Aligned: A powerful Predictor of Stock Returns. *Review of Financial Studies*, *28*(3), 791–837.

Jagarlamudi, J., Daumé III, H., & Udupa, R. (2012). Incorporating Lexical Priors into Topic Models. *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, 204–213.

Jiang, F., Lee, J., Martin, X., & Zhou, G. (2019). Manager Sentiment and Stock Returns. *Journal of Financial Economics*, *132*(1), 126–149.

Julliard, C., & Ghosh, A. (2012). Can Rare Events Explain the Equity Premium Puzzle? *Review of Financial Studies*, *25*(10), 3037–3076.

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

Kelly, B., & Pruitt, S. (2013). Market Expectations in the Cross-Section of Present Values. *Journal of Finance*, *68*(5), 1721–1756.

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

Larsen, V. H., & Thorsrud, L. A. (2019). The Value of News for Economic Developments. *Journal of Econometrics*, *210*(1), 203–218.

Le Bris, D. (2012). Wars, inflation and stock market returns in france, 1870–1945. *Financial History Review*, *19*(3), 337–361.

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

Loughran, T., & McDonald, B. (2016). Textual Analysis in Accounting and Finance: A Survey. *Journal of Accounting Research*, *54*(4), 1187–1230.

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

Lu, B., Ott, M., Cardie, C., & Tsou, B. K. (2011). Multi-Aspect Sentiment Analysis with Topic Models. *2011 IEEE 11th International Conference on Data Mining Workshops*, 81–88.

Lundblad, C. (2007). The Risk Return Tradeoff in the Long Run: 1836–2003. *Journal of Financial Economics*, *85*(1), 123–150.

Manela, A., & Moreira, A. (2017). News Implied Volatility and Disaster Concerns. *Journal of Financial Economics*, *123*(1), 137–162.

Mcauliffe, J., & Blei, D. (2007). Supervised Topic Models. *Advances in Neural Information Processing Systems*, *20*, 121–128.

Merton, R. C. (1973). An Intertemporal Capital Asset Pricing Model. *Econometrica*, *41*(5), 867–887.

Nelson, D. B. (1991). Conditional Heteroskedasticity in Asset Returns: A New Approach. *Econometrica*, *59*(2), 347–370.

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

Obaid, K., & Pukthuanthong, K. (2022). A Picture is Worth a Thousand Words: Measuring Investor Sentiment by Combining Machine Learning and Photos from News. *Journal of Financial Economics*, *144*(1), 273–297.

Oosterlinck, K., & Landon-Lane, J. S. (2006). Hope Springs Eternal–French Bondholders and the Soviet Repudiation (1915–1919). *Review of Finance*, *10*(4), 507–535.

Pástor, L., & Veronesi, P. (2013). Political Uncertainty and Risk Premia. *Journal of Financial Economics*, *110*(3), 520–545.

Ramage, D., Hall, D., Nallapati, R., & Manning, C. D. (2009). Labeled lda: A supervised topic model for credit attribution in multi-labeled corpora. *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, 248–256.

Rapach, D. E., Ringgenberg, M. C., & Zhou, G. (2016). Short Interest and Aggregate Stock Returns. *Journal of Financial Economics*, *121*(1), 46–65.

Rapach, D. E., Strauss, J. K., & Zhou, G. (2010). Out-of-Sample Equity Premium Prediction: Combination Forecasts and Links to the Real Economy. *The Review of Financial Studies*, *23*(2), 821–862.

Rhodes, R. (2012). *The Making of the Atomic Bomb*. Simon; Schuster.

Rietz, T. A. (1988). The Equity Risk Premium: A Solution? *Journal of Monetary Economics*, *22*(1), 117–131.

Rutherford, E. (2012). The Scattering of α and β Particles by Matter and the Structure of the Atom. *Philosophical Magazine*, *92*(4), 379–398.

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

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.

Snowberg, E., & Wolfers, J. (2010). Explaining the Favorite–Long Shot Bias: Is It Risk-Love or Misperceptions? *Journal of Political Economy*, *118*(4), 723–746.

Stambaugh, R. F., Yu, J., & Yuan, Y. (2012). The Short of It: Investor Sentiment and Anomalies. *Journal of Financial Economics*, *104*(2), 288–302.

Steyvers, M., & Griffiths, T. (2007). Probabilistic topic models. In *Handbook of latent semantic analysis* (pp. 439–460). Psychology Press.

Tetlock, P. C. (2007). Giving Content to Investor Sentiment: The Role of Media in the Stock Market. *Journal of Finance*, *62*(3), 1139–1168.

Tversky, A., & Kahneman, D. (1992). Advances in Prospect Theory: Cumulative Representation of Uncertainty. *Journal of Risk and Uncertainty*, *5*, 297–323.

van Binsbergen, J. H., Bryzgalova, S., Mukhopadhyay, M., & Sharma, V. (2022). (Almost) 200 Years of News-Based Economic Sentiment. *Available at SSRN 4261249*.

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

Walker, J. S. (2004). *Three Mile Island: A Nuclear Crisis in Historical Perspective*. University of California Press.

Watanabe, K., & Zhou, Y. (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.

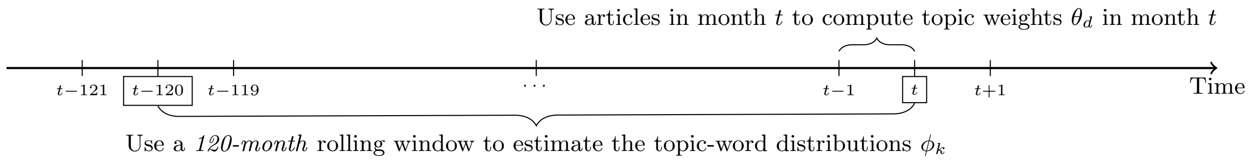
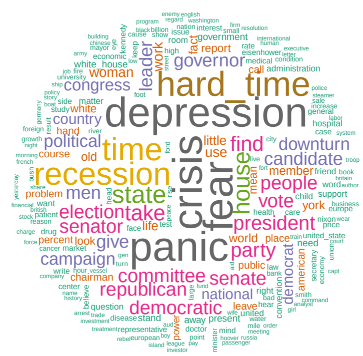


Figure 2. Discourse Topic Contents

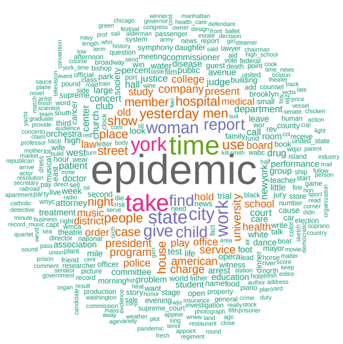
This figure plots the frequencies of n-grams per topic over time. Frequencies are constructed according to the sLDA model described in Section 2, and the size of each n-gram indicates its frequency. The sample period is from January 1871 to October 2019.

War

Panic

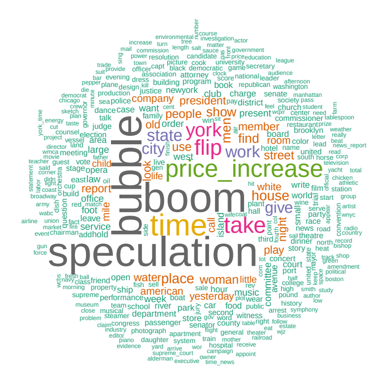


Pandemic



Monetary Standard

Real Estate Booms



Boycotts and Evil Business

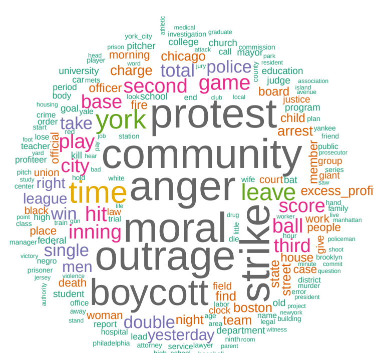


Figure 3. Time Series of Discourse Topic Weights

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.

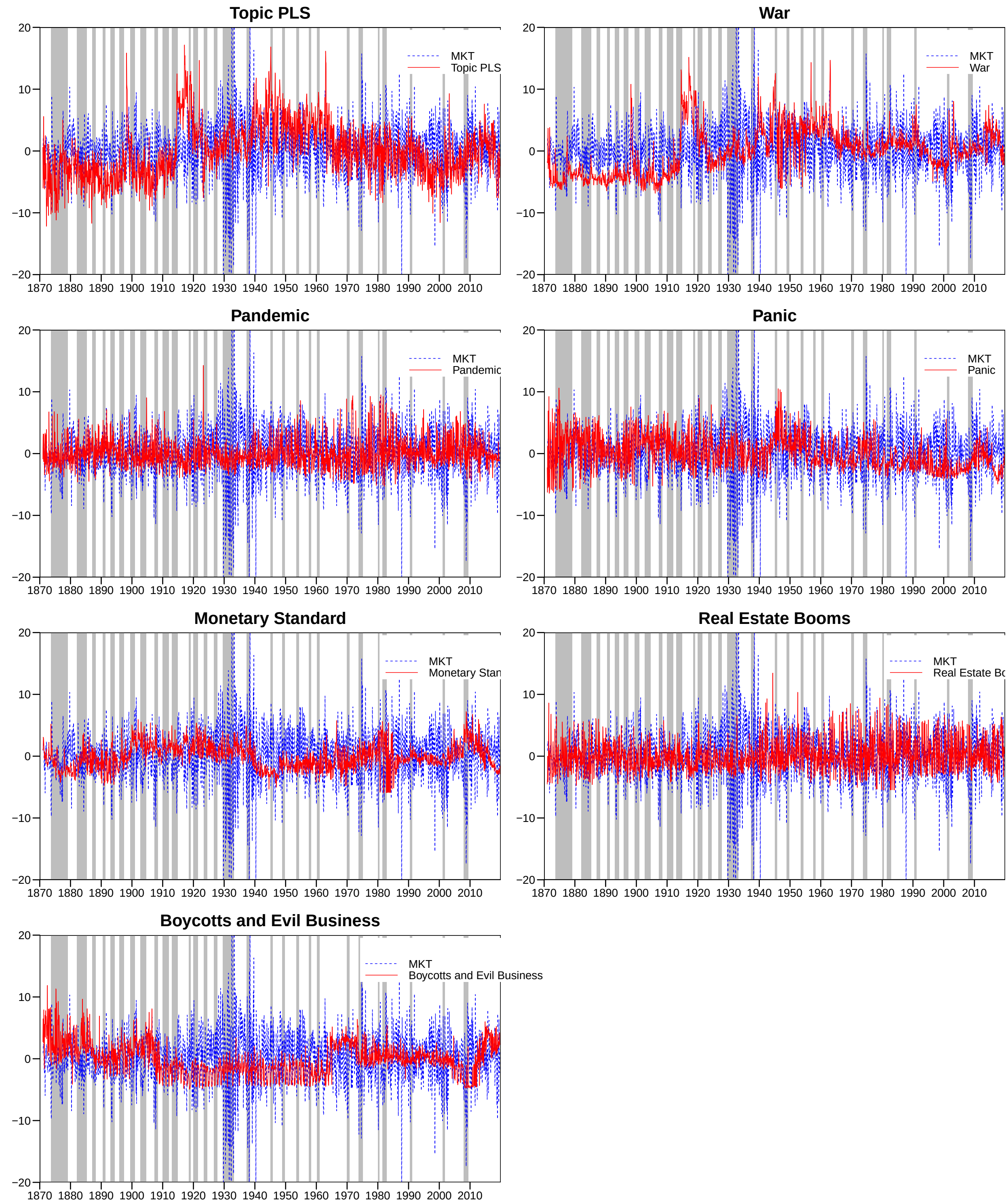


Figure 4. Articles Making the Biggest Contribution to *War* Spikes since 1990

This figure plots the ten articles that have contributed significantly to ten monthly heights of *War* since 1990. Topic weights are demeaned. The gray-shaded areas represent NBER-defined recessions. The sample period is from January 1990 to October 2019.

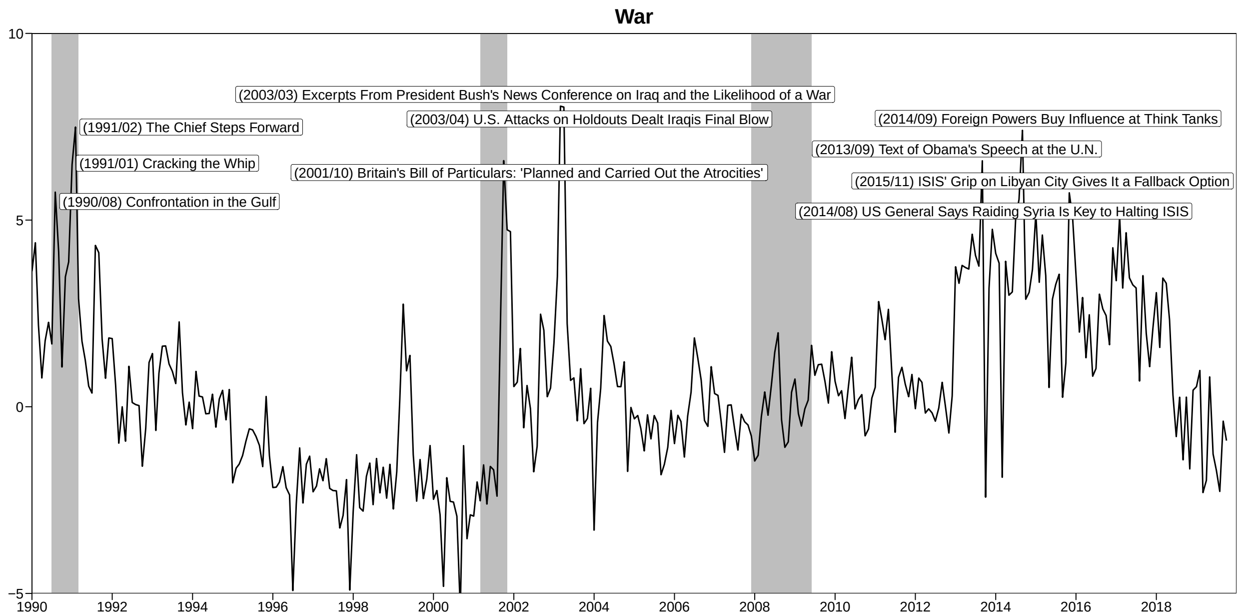


Figure 5. Cumulative R-Squared in One-Month Return Prediction

Panel A plots the cumulative in-sample R^2 computed as

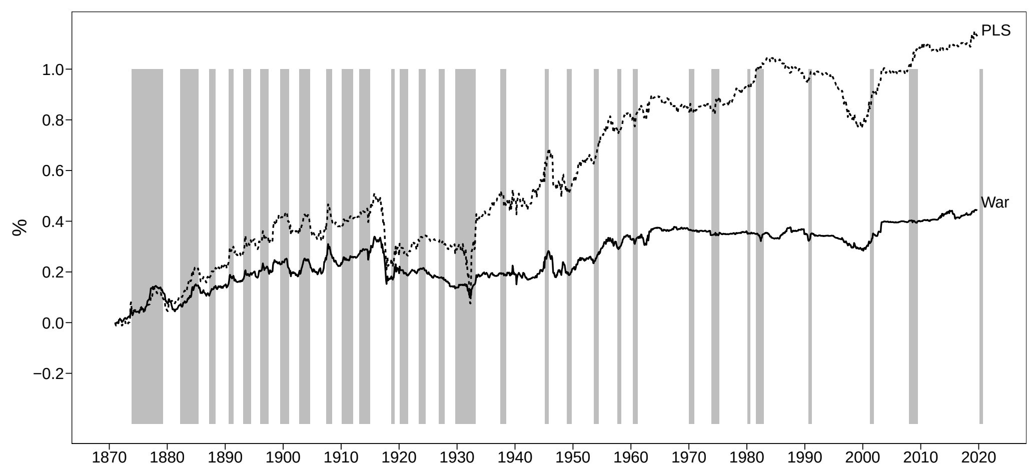
$$\left(\sum_{s=1}^{t}\left(R_s^e-\bar{R}^e\right)^2-\sum_{s=1}^{t}\left(R_s^e-\hat{R}_s^e\right)^2\right)/\sum_{s=1}^{T}\left(R_s^e-\bar{R}^e\right)^2,$$

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. Panel B 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 the historical mean and predicted value, estimated based on the preceding estimation window. The evaluation period is from January 1881 to October 2019.

Panel A: In-Sample R^2



Panel B: Out-of-Sample R^2

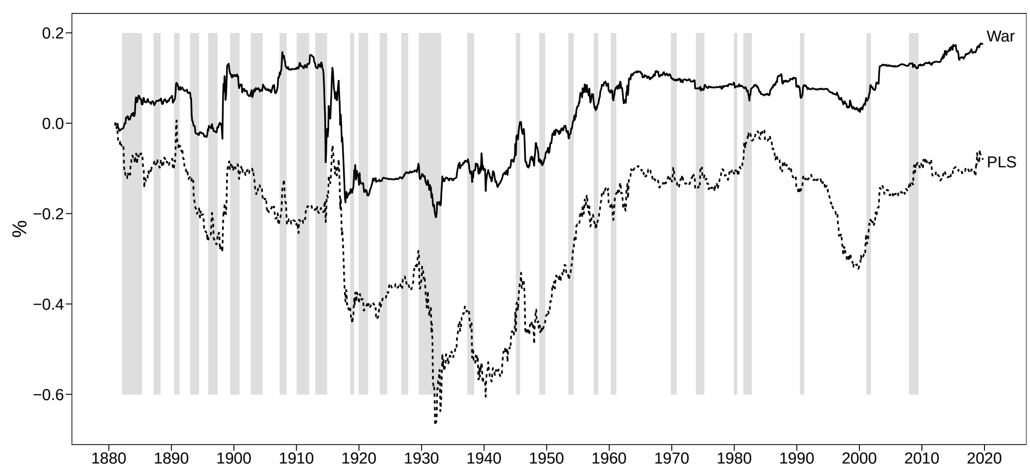


Table 1
Seed Words

This table lists the *lemmatized* seed words for each of the 14 discourse topics. The first column presents the full name of the topic, and the second column reports the short name used in the paper.

Narrative	Short Name	Seed Words
War	War	conflict, tension, terrorism, war
Pandemic	Pandemic	epidemic, pandemic
Panic	Panic	bank failure, bank panic, bank run, crisis, depression, downturn, fear, financial panic, hard time, panic, recession
Confidence	Confidence	business confidence, consumer confidence
Frugality	Saving	compassion, family morale, frugal, frugality, modesty, moral, poverty, saving
Conspicuous Consumption	Consumption	american dream, conspicuous consumption, consumption, equal opportunity, equality, homeownership, luxury, patriotism, prosperity
Monetary Standard	Money	bimetallism, devaluation, gold, gold standard, inflation, monetary standard, money, silver
Techmology Replacing Jobs	Tech	automate, computer, digital divide, electronic brain, invention, labor save, labor save machine, machine, mechanize, network, technocracy, technological unemployment, technology, unemployment
Real Estate Booms	Real_estate_boom	boom, bubble, flip, flipper, home ownership, home purchase, house boom, house bubble, land boom, land bubble, price increase, real estate boom, real estate bubble, speculation
Real Estate Crashes	Real_estate_crash	bust, crash, house bust, house crash, land bust, land crash, price decrease, real estate bust, real estate crash
Stock Market Bubbles	Stock_bubble	advance market, boom, bubble, bull, bull market, bullish, earnings per share, inflate market, margin, margin requirement, market boom, market bubble, price earn ratio, price increase, sell short, short sell, speculation, stock market boom, stock market bubble
Stock Market Crashes	Stock_crash	bear, bear market, bearish, bust, crash, fall market, market crash, stock crash, 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 14 monthly topic weights constructed according to the sLDA model described in Section 2. All numbers (except sample size) are expressed as percentages. The sample period is from January 1871 to October 2019.

	N	Mean	SD	Q1	Median	Q3	AC(1)	PLS Weights	Corr PLS
War	1784	9.71	3.73	6.56	9.64	11.92	84.84	5.22	82.09
Pandemic	1784	5.72	2.34	4.25	5.38	6.72	7.54	-2.26	-28.53
Panic	1784	8.30	2.70	6.21	7.94	10.13	37.03	2.19	15.74
Confidence	1784	5.72	2.45	3.97	5.39	6.99	7.61	0.60	-0.47
Saving	1784	5.84	2.17	4.39	5.51	6.84	29.78	-1.09	-34.12
Consumption	1784	7.36	2.85	5.46	6.72	9.23	27.62	0.88	-4.58
Money	1784	6.58	2.06	5.21	6.46	7.87	60.55	-1.77	-15.26
Tech	1784	6.61	2.52	4.99	6.57	8.07	54.58	-0.79	-8.15
Real estate boom	1784	5.95	2.52	4.20	5.60	7.22	9.31	-2.82	-32.73
Real estate crash	1784	5.57	2.16	4.23	5.41	6.49	23.16	0.47	-1.12
Stock bubble	1784	5.79	2.30	4.28	5.74	7.23	48.98	-0.40	-14.67
Stock crash	1784	7.40	3.13	5.06	6.86	9.60	26.56	0.86	-2.31
Boycott	1784	5.79	2.62	4.14	5.51	7.37	67.00	-1.47	-41.69
Wage	1784	7.90	2.59	6.05	7.60	9.50	38.53	1.07	37.33
PLS	1784	49.99	44.73	18.75	46.68	77.28	70.35		

Table 3
Predicting One-Month Market Returns

This table presents the results of the following predictive regression:

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

where R^e_{t+1} is the excess market return over the next month, x_t is one of the discourse topics or the PLS indexes, and β, the coefficient of interest, measures the strength of predictability. "Shiller PLS" uses only the topics from Shiller (2019), excluding *War* and Pandemic. 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 January 1871 to October 2019. *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

	1871-2019	1871-1949	1950-2019	2000-2019
War (%)	3.80 ***	3.49 **	4.06 **	9.83 ***
t-stat	(3.35)	(2.02)	(2.56)	(3.43)
R^2 (%)	0.39	0.20	0.55	3.39
Pandemic (%)	-2.61 **	-3.98 **	-1.55	-2.31
t-stat	(-2.06)	(-2.14)	(-0.91)	(-0.70)
R^2 (%)	0.15	0.29	-0.02	-0.21
Panic (%)	2.21 *	3.06 *	2.57 *	3.36
t-stat	(1.76)	(1.71)	(1.69)	(1.20)
R^2 (%)	0.09	0.13	0.15	0.02
Confidence (%)	0.66	-0.14	1.13	0.77
t-stat	(0.55)	(-0.08)	(0.67)	(0.24)
R^2 (%)	-0.04	-0.11	-0.07	-0.40
Saving (%)	-1.37	-1.62	-1.44	-0.98
t-stat	(-1.05)	(-0.92)	(-0.81)	(-0.37)
R^2 (%)	0.00	-0.04	-0.04	-0.39
Consumption (%)	0.84	0.63	2.58	0.07
t-stat	(0.66)	(0.32)	(1.61)	(0.02)
R^2 (%)	-0.03	-0.10	0.15	-0.42
Money (%)	-2.33	-1.23	-3.41 *	-0.70
t-stat	(-1.64)	(-0.61)	(-1.76)	(-0.17)
R^2 (%)	0.11	-0.07	0.36	-0.40
Tech (%)	-0.85	0.54	-3.10	-14.25 ***
t-stat	(-0.58)	(0.26)	(-1.62)	(-3.24)
R^2 (%)	-0.03	-0.10	0.27	7.59
Real estate boom (%)	-3.03 **	-3.51 **	-3.02 *	-6.57 **
t-stat	(-2.50)	(-2.05)	(-1.82)	(-2.16)
R^2 (%)	0.23	0.20	0.25	1.28
Real estate crash (%)	0.59	0.59	0.93	-2.58
t-stat	(0.48)	(0.34)	(0.55)	(-0.78)
R^2 (%)	-0.05	-0.10	-0.08	-0.16
Stock bubble (%)	-0.47	-1.67	0.39	-4.59
t-stat	(-0.37)	(-0.96)	(0.21)	(-1.31)
R^2 (%)	-0.05	-0.04	-0.11	0.41
Stock crash (%)	0.75	2.08	-0.23	-0.83
t-stat	(0.54)	(1.02)	(-0.14)	(-0.26)
R^2 (%)	-0.04	0.00	-0.12	-0.40
Boycott (%)	-1.52	-2.36	-0.43	5.33
t-stat	(-1.23)	(-1.41)	(-0.23)	(1.38)
R^2 (%)	0.01	0.04	-0.11	0.70
Wage (%)	1.12	0.70	1.87	8.62 ***
t-stat	(0.74)	(0.32)	(1.04)	(2.81)
R^2 (%)	-0.02	-0.09	0.02	2.51
PLS (%)	6.07 ***	5.86 ***	6.43 ***	12.17 ***
t-stat	(4.59)	(2.98)	(3.93)	(4.33)
R^2 (%)	1.07	0.76	1.57	5.42
Shiller PLS (%)	4.76 ***	5.98 ***	5.37 ***	7.46 **
t-stat	(3.15)	(2.71)	(3.23)	(2.30)
R^2 (%)	0.64	0.80	1.06	1.77

Table 4
Predicting Long-Horizon Market Returns

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 either *War* or the PLS indexes constructed from 14 topics, and β, the coefficient of interest, measures the strength of predictability. "Expanding PLS" is the PLS index recursively estimated every month using only the data available up to that month. 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 1871 to October 2019. *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

	War (%)	t-stat	R^2 (%)	PLS (%)	t-stat	R^2 (%)	Expanding PLS (%)	t-stat	R^2 (%)	N
Panel A: 1871-2019										
$h = 1$	3.80 ***	(3.35)	0.39	6.07 ***	(4.59)	1.07	4.65 ***	(3.52)	0.58	1664
$h = 3$	2.87 ***	(2.81)	0.57	4.20 ***	(3.88)	1.29	3.72 ***	(3.55)	0.96	1664
$h = 6$	3.03 ***	(2.83)	1.36	3.69 ***	(3.58)	2.03	3.28 ***	(3.26)	1.56	1664
$h = 12$	2.79 **	(2.48)	1.93	3.11 ***	(2.92)	2.42	2.76 ***	(2.66)	1.86	1664
$h = 24$	2.09 **	(2.08)	1.91	2.25 **	(2.30)	2.22	2.39 **	(2.39)	2.55	1656
$h = 36$	2.28 **	(2.22)	3.07	2.63 ***	(2.59)	4.12	3.06 ***	(3.10)	5.72	1644
Panel B: 1871-1949										
$h = 1$	3.49 **	(2.02)	0.20	5.86 ***	(2.98)	0.76	3.92 *	(1.82)	0.24	828
$h = 3$	2.51	(1.60)	0.26	4.01 **	(2.44)	0.83	3.00 *	(1.75)	0.37	828
$h = 6$	2.88 *	(1.77)	0.94	3.71 **	(2.39)	1.63	3.00 *	(1.85)	0.95	828
$h = 12$	2.87 *	(1.73)	1.59	3.47 **	(2.23)	2.36	2.76 *	(1.74)	1.36	828
$h = 24$	1.90	(1.38)	1.25	2.55 *	(1.94)	2.32	2.54 *	(1.73)	2.27	828
$h = 36$	2.11	(1.45)	2.17	2.90 **	(2.09)	4.20	3.32 **	(2.31)	5.63	828
Panel C: 1950-2019										
$h = 1$	4.06 **	(2.56)	0.55	6.43 ***	(3.93)	1.57	5.33 ***	(3.35)	1.04	836
$h = 3$	3.18 **	(2.41)	1.06	4.50 ***	(3.78)	2.24	4.40 ***	(3.64)	2.14	836
$h = 6$	2.93 **	(2.24)	1.64	3.61 ***	(3.13)	2.55	3.41 ***	(2.83)	2.26	836
$h = 12$	2.21	(1.50)	1.64	2.49 *	(1.92)	2.11	2.56 *	(1.94)	2.23	836
$h = 24$	1.85	(1.24)	1.93	1.66	(1.21)	1.53	1.99	(1.54)	2.25	828
$h = 36$	1.96	(1.36)	2.85	2.09	(1.55)	3.27	2.49 *	(1.90)	4.70	816
Panel D: 2000-2019										
$h = 1$	9.83 ***	(3.43)	3.39	12.17 ***	(4.33)	5.42	9.61 ***	(3.58)	3.22	238
$h = 3$	7.64 ***	(4.07)	6.07	7.25 ***	(4.25)	5.42	5.24 ***	(3.12)	2.63	238
$h = 6$	6.08 ***	(3.31)	6.76	6.35 ***	(4.33)	7.40	4.05 ***	(3.13)	2.76	238
$h = 12$	5.42 **	(2.47)	9.68	5.21 ***	(3.28)	8.91	3.10 **	(2.51)	2.89	238
$h = 24$	4.78 **	(2.28)	12.78	4.35 ***	(2.66)	10.51	2.47 *	(1.96)	3.11	230
$h = 36$	4.59 ***	(2.86)	17.72	4.83 ***	(3.88)	19.64	3.24 ***	(3.43)	8.61	218

Table 5
Predicting One-Month Market Returns: *War* versus Other Predictors

This table presents the results of the following predictive regressions:

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

where R^e_{t+1} is the excess market return over the next month and z_t is one of the economic predictors from Goyal and Welch (2008) (Panel A) or the remaining topics (Panel B). The last column reports the results when *War* is tested against all predictors in each panel. 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 January 1871 to October 2019. *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

Panel A: Economic Predictors

	(1)	(2)	(3)	(4)	(5)	(6)
War	3.85 ***	3.55 ***	3.75 ***	3.80 ***	3.23 ***	2.52 **
	(3.36)	(3.08)	(2.97)	(3.36)	(2.81)	(1.98)
DP	1.54					-0.90
	(0.80)					(-0.34)
EP		2.03				3.88
		(1.29)				(1.44)
DE			-0.25			
			(-0.11)			
SVAR				-0.17		-0.28
				(-0.04)		(-0.07)
TBL					-3.09 *	-4.21 ***
					(-1.92)	(-2.76)
R^2	0.40	0.46	0.33	0.33	0.61	0.75

Table 5
Predicting One-Month Market Returns: *War* versus Other Predictors (Cont.)

Panel B: Other Topic Predictors

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
War	3.64 ***	4.30 ***	3.87 ***	3.67 ***	4.31 ***	3.60 ***	3.78 ***	3.74 ***	3.89 ***	3.79 ***	4.11 ***	3.69 ***	3.73 ***	5.38 **
	(3.21)	(3.74)	(3.41)	(3.13)	(3.51)	(3.16)	(3.35)	(3.30)	(3.39)	(3.34)	(3.48)	(3.08)	(3.20)	(2.29)
Pandemic	-2.38 *													-1.48
	(-1.88)													(-0.89)
Panic		2.94 **												3.33 *
		(2.31)												(1.67)
Confidence			0.97											1.63
			(0.81)											(0.89)
Saving				-0.55										0.48
				(-0.41)										(0.28)
Consumption					1.97									2.12
					(1.44)									(1.13)
Money						-1.97								-1.82
						(-1.37)								(-1.04)
Tech							-0.79							-0.29
							(-0.55)							(-0.15)
Real_estate_boom								-2.97 **						-2.05
								(-2.45)						(-1.24)
Real_estate_bust									0.98					1.33
									(0.79)					(0.80)
Stock_bubble										-0.09				0.72
										(-0.07)				(0.41)
Stock_crash											1.56			2.18
											(1.10)			(0.96)
Boycott												-0.33		0.11
												(-0.25)		(0.06)
Wage													0.27	0.55
													(0.18)	(0.26)
R^2	0.50	0.59	0.36	0.34	0.44	0.45	0.35	0.60	0.36	0.33	0.40	0.33	0.33	0.67

Table 6
Predicting One-Month Market Returns:
***War* versus Other Media-Based Uncertainty and Crisis Event Count Indexes**

This table presents the results of the following bivariate predictive regressions:

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

where R^e_{t+1} is the excess market return over the next month and x_t is *War*. In Panel A, z_t is either NVIX^2 from Manela and Moreira (2017), or geopolitical risk (GPR) from Caldara and Iacoviello (2022); in Panel B, z_t is either Crisis (monthly count of real-word crisis events), or CWar (monthly count of real-word war events) from Berkman, Jacobsen, and Lee (2011). Returns are expressed as annualized percentages, and the independent variables are 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. *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$. The whole sample in Panel A is from January 1900 to March 2016 and in Panel B is from January 1918 to December 2018.

Panel A: Other Media-Based Uncertainty Indexes

			1900-2016		
War	2.69 **			2.70 **	
	(1.99)			(2.00)	
NVIX^2		0.07		0.22	0.07
		(0.03)		(0.10)	(0.03)
GPR			2.48 *		2.48 *
			(1.72)		(1.72)
R^2	0.12	-0.07	0.09	0.05	0.02
			1950-2016		
War	3.96 **			3.90 **	
	(2.42)			(2.29)	
NVIX^2		1.01		0.67	0.96
		(0.32)		(0.21)	(0.31)
GPR			3.51 *		3.50 *
			(1.91)		(1.91)
R^2	0.50	-0.09	0.37	0.39	0.28
			2000-2016		
War	10.08 ***			10.19 ***	
	(3.11)			(3.05)	
NVIX^2		-0.63		-1.42	-0.07
		(-0.11)		(-0.25)	(-0.01)
GPR			7.32 **		7.32 **
			(2.27)		(2.32)
R^2	3.19	-0.50	1.44	2.75	0.92

Table 6
Predicting One-Month Market Returns:
***War* versus Other Media-Based Uncertainty and Crisis Event Count Indexes**

Panel B: Crisis Event Count Indexes

		1918-2018		
War	3.09 **			2.66 *
	(2.08)			(1.75)
Crisis		1.60		-0.36
		(0.99)		(-0.20)
CWar			3.75 **	3.52 *
			(1.99)	(1.69)
R^2	0.15	-0.02	0.26	0.27
		1950-2018		
War	4.12 **			4.28 ***
	(2.51)			(2.58)
Crisis		1.01		0.11
		(0.66)		(0.07)
CWar			1.58	1.93
			(0.97)	(1.13)
R^2	0.57	-0.08	-0.02	0.48
		2000:2018		
War	10.23 ***			10.05 ***
	(3.30)			(2.97)
Crisis		-2.59		-2.00
		(-0.92)		(-0.64)
CWar			-3.42	-0.37
			(-1.17)	(-0.11)
R^2	3.64	-0.18	0.01	2.96

Table 7
Predicting One-Month Market Returns after Controlling for Economic, Sentiment, and Uncertainty Predictors

This table presents the results of the following univariate predictive regression:

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

and the following bivariate predictive regression:

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

where R^e_{t+1} is the excess market return over the next month and x_t is the discourse topic PLS index. In Panel A, z_t is one of the 14 economic predictors from Goyal and Welch (2008), output gap from Cooper and Priestley (2009), and short interest from Rapach, Ringgenberg, and Zhou (2016); in Panel B, z_t is one of the uncertainty variables (financial and macro uncertainty indexes from Jurado, Ludvigson, and Ng (2015), economic policy uncertainty index from Baker, Bloom, and Davis (2016), disagreement index from Huang, Li, and Wang (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 of each panel reports the results using the PLS index constructed from all predictors in that panel. 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: Economic Predictors

	Univariate		Bivariate			
Economic Predictor	γ(%)	R^2(%)	β(%)	γ(%)	R^2(%)	Period
Dividend-price ratio (DP)	1.39	0.00	6.00 ***	0.40	1.02	187101-201910
Dividend yield (DY)	2.03	0.07	5.89 ***	1.03	1.05	187102-201910
Earnings-price ratio (EP)	2.46	0.13	5.80 ***	1.23	1.06	187101-201910
Dividend payout ratio (DE)	-1.00	-0.03	6.05 ***	-0.87	1.04	187101-201910
Stock variance (SVAR)	-0.08	-0.06	6.08 ***	-0.37	1.02	187101-201910
Book-to-market Ratio (BM)	5.19	0.58	5.09 ***	3.49	1.06	192103-201910
Net equity expansion (NTIS)	-4.11	0.31	6.87 ***	-4.83	1.33	192612-201910
Treasury bill rate (TBL)	-3.68 **	0.36	5.46 ***	-2.11	1.14	187101-201910
Long term bond yield (LTY)	-2.82	0.11	6.25 ***	-0.60	0.87	191901-201910
Long term bond return (LTR)	3.36 *	0.18	6.62 ***	3.88 **	1.13	192601-201910
Term spread (TMS)	-2.70	0.10	6.30 ***	-0.46	0.87	191901-201910
Default yield spread (DFY)	2.87	0.12	6.37 ***	2.65	1.03	191901-201910
Default return spread (DFR)	2.30	0.04	6.29 ***	2.21	0.89	192601-201910
Inflation (INFL)	-3.32	0.20	5.70 ***	-3.60	0.95	191302-201910
Output Gap (OG)	-3.39	0.20	6.33 ***	-3.24	1.10	191902-201910
Short Interest (SI)	-5.70 **	0.94	4.98 **	-5.27 **	1.60	197301-201412
Economic PLS	4.40 *	0.48	4.53 *	3.05	0.93	197301-201412

Table 7
Predicting One-Month Market Returns after
Controlling for Economic, Sentiment, and Uncertainty Predictors (Cont.)

Panel B: Uncertainty and Sentiment Predictors

	Correlations	Univariate		Bivariate			
Economic Predictor	Corr. with PLS (%)	γ(%)	R^2(%)	β(%)	γ(%)	R^2(%)	Period
Financial uncertainty	-5.77	-5.75 **	1.15	4.95 ***	-5.47 *	1.96	196007-201910
Macro uncertainty	-2.43	-4.30	0.58	5.17 ***	-4.17	1.48	196007-201910
Economic policy uncertainty	25.51 ***	4.03	0.38	4.02 *	3.01	0.73	198501-201910
Implied volatility (VIX)	-1.62	0.40	-0.27	6.32 **	0.50	1.11	199001-201910
News implied volatility (NVIX)	6.55 **	0.03	-0.07	5.80 ***	-0.35	0.80	188907-201603
Disagreement	-8.12 **	-8.43 ***	2.43	4.98 ***	-8.02 ***	3.17	196912-201812
News sentiment	14.71 ***	-0.52	-0.05	6.28 ***	-1.44	1.08	186612-201910
Investor sentiment (BW)	-13.83 ***	-2.50	0.08	4.68 **	-1.86	0.73	196507-201812
Investor sentiment (PLS)	-2.54	-7.32 ***	1.86	4.75 **	-7.20 ***	2.56	196507-201812
Manager sentiment	-19.31 ***	-9.06 **	3.32	6.87 ***	-7.74 **	4.93	200301-201712
Shiller's one-year confidence index	-29.87 ***	-4.77	0.48	10.48 ***	-1.64	4.17	200107-201910
Shiller's crash confidence index	-22.47 ***	-2.07	-0.28	11.06 ***	0.41	4.07	200107-201910
Uncertainty PLS	18.43 **	2.38	-0.39	8.84 ***	0.75	2.22	200301-201603

Table 8
Out-of-Sample R^2

This table reports the out-of-sample R^2 (R^2_{OS}) statistic (Campbell and Thompson 2008) in predicting the monthly excess market return using economic predictors or discourse topics. "Shiller Topics" uses only the topics from Shiller (2019), excluding *War* and Pandemic. Panels A and B report individual OLS regressions and the PLS method results, 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 1881, 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).

	1881-2019	1881-1949	1950-2019	2000-2019
Panel A: OLS				
Dividend-price ratio (DP)	-0.60	-0.81	-0.25	0.05
Dividend yield (DY)	-0.48	-0.39	-0.64	0.04
Earnings-price ratio (EP)	-0.14	-0.07	-0.26	-0.35
Dividend payout ratio (DE)	-0.83	-1.12	-0.33	-1.06
Stock variance (SVAR)	-1.68	-2.18	-0.79	-0.86
Treasury bill rate (TBL)	0.07 **	-0.05	0.26 **	0.45
War	0.17 ***	-0.10 **	0.65 ***	1.35 ***
Pandemic	0.08 *	0.27 **	-0.23	0.18
Panic	0.04	0.11	-0.06	0.28
Confidence	-0.09	-0.11	-0.07	-0.09
Saving	-0.03	-0.05	-0.00	0.02
Consumption	-0.10	-0.14	-0.03	-0.01
Money	0.01	-0.18	0.33 *	-0.19
Tech	-0.45	-0.69	-0.01	0.12
Real estate boom	0.19 **	0.21 *	0.14 *	1.13 **
Real estate crash	-0.11	-0.15	-0.03	-0.15
Stock bubble	-0.12	-0.05	-0.23	-0.27
Stock crash	-0.10	-0.03	-0.23	-0.07
Boycott	-0.03	0.02	-0.11	-1.24
Wage	-0.16	-0.27	0.03	0.32 **
Panel B: PLS				
Economic	-0.84	-1.11	-0.38	-0.55
Topics	-0.08 ***	-0.67	0.95 ***	2.23 ***
Shiller Topics	-0.88	-1.03	-0.62	-0.17

Table 9
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 14 discourse topics to make return forecasts compared to historical mean returns. "Shiller Topics" uses only the topics from Shiller (2019), excluding *War* and Pandemic. Panels A and B report the results using OLS and PLS, respectively. The last row reports the annualized monthly Sharp ratio of the S&P 500 index. All out-of-sample forecasts are estimated recursively using the data available in the expanding estimation window. The evaluation period begins in January 1881, 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, Garlappi, and Uppal (2009)).

	Utility Gain (%)				Sharpe Ratio			
	1881-2019	1881-1949	1950-2019	2000-2019	1881-2019	1881-1949	1950-2019	2000-2019
Panel A: OLS								
Dividend-price ratio (DP)	-0.31	0.07	-0.70	0.24	0.36	0.25	0.48	0.21
Dividend yield (DY)	-0.50	-0.08	-0.94	1.02	0.35	0.23	0.50	0.30
Earnings-price ratio (EP)	0.59	0.53	0.64	3.88 *	0.44	0.28	0.56	0.65 ***
Dividend payout ratio (DE)	0.42	0.92	-0.08	-0.10	0.42	0.32 *	0.49	0.26
Stock variance (SVAR)	-0.44	-0.36	-0.53	-0.44	0.35	0.19	0.46	0.23
Treasury bill rate (TBL)	1.31 **	0.89	1.72 *	1.93 *	0.48 **	0.32 **	0.62 **	0.40 **
War	0.86 **	0.87	0.85 **	2.01 ***	0.45 **	0.32 *	0.55 **	0.38 ***
Pandemic	-0.09	-0.19	0.02	1.18	0.38	0.21	0.50	0.32
Panic	0.37	1.04 *	-0.31	-0.18	0.42	0.33 *	0.49	0.21
Confidence	-0.02	0.01	-0.05	-0.13	0.38	0.23	0.49	0.25
Saving	-0.04	-0.09	0.01	-0.21	0.38	0.21	0.50	0.24
Consumption	0.08	0.13	0.03	-0.04	0.39	0.24	0.50	0.24
Money	1.03 ***	0.66	1.40 **	1.86	0.47 ***	0.30	0.59 **	0.36
Tech	-0.53	-1.09	0.04	0.33	0.34	0.13	0.50	0.27
Real estate boom	0.24	0.15	0.32	1.89	0.40	0.24	0.52	0.37
Real estate crash	-0.06	-0.17	0.05	-0.08	0.38	0.20	0.50	0.25
Stock bubble	-0.14	0.06	-0.34	-0.25	0.37	0.23	0.47	0.24
Stock crash	-0.09	0.16	-0.35	-0.03	0.38	0.24	0.47	0.25
Boycott	0.20	0.02	0.38	-0.73	0.40	0.23	0.52	0.22
Wage	-0.17	-0.31	-0.04	0.10	0.37	0.19	0.49	0.26
Panel B: PLS								
Economic	0.44	0.70	0.16	3.08	0.42	0.31	0.53	0.65 **
Topics	0.69	0.66	0.70	4.11 **	0.43	0.31 *	0.55	0.54 **
Shiller Topics	0.46	0.50	0.40	1.05	0.43	0.29	0.56	0.29
Buy and Hold					0.39	0.28	0.55	0.35

Table 10
Predicting Bond Returns

This table presents the results of the following predictive regression:

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

where $R^e_{t+1\to t+h}$ is the excess returns over the next h months on Datastream Treasury bond indexes (Panel A), S&P investment grade corporate bond (Panel B), and S&P high yield corporate bond (Panel C). 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 is from December 1988 to October 2019. *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

Panel A: Government Bond Indexes

	h=1	h=3	h=6	h=12
	1988-2019			
U.S. Goverment Bond 1-3 Years	-0.290	-0.225	-0.192	-0.133
(t-stat)	(-1.41)	(-1.13)	(-1.00)	(-0.52)
R^2	0.12	0.25	0.41	0.22
U.S. Goverment Bond 3-5 Years	-0.57	-0.44	-0.33	-0.16
(t-stat)	(-1.03)	(-0.99)	(-0.94)	(-0.38)
R^2	-0.04	0.07	0.11	-0.11
U.S. Goverment Bond 5-7 Years	-0.73	-0.60	-0.42	-0.16
(t-stat)	(-0.91)	(-0.98)	(-0.88)	(-0.33)
R^2	-0.07	0.09	0.07	-0.17
U.S. Goverment Bond 7-10 Years	-0.91	-0.81	-0.50	-0.13
(t-stat)	(-0.86)	(-1.01)	(-0.80)	(-0.24)
R^2	-0.09	0.14	0.03	-0.23
U.S. Goverment Bond 10+ Years	-0.94	-0.78	-0.22	0.16
(t-stat)	(-0.54)	(-0.57)	(-0.20)	(0.21)
R^2	-0.20	-0.13	-0.25	-0.25
	2000-2019			
U.S. Goverment Bond 1-3 Years	-0.591 **	-0.529 **	-0.586 ***	-0.590 **
(t-stat)	(-2.43)	(-2.41)	(-2.64)	(-2.21)
R^2	1.58	3.30	7.07	10.98
U.S. Goverment Bond 3-5 Years	-1.07	-0.89 *	-0.93 **	-0.88 **
(t-stat)	(-1.53)	(-1.86)	(-2.39)	(-2.08)
R^2	0.44	1.26	2.98	5.92
U.S. Goverment Bond 5-7 Years	-1.31	-1.11	-1.06 *	-0.91 *
(t-stat)	(-1.27)	(-1.60)	(-1.95)	(-1.77)
R^2	0.20	0.93	1.95	3.60
U.S. Goverment Bond 7-10 Years	-1.52	-1.31	-1.07	-0.77
(t-stat)	(-1.09)	(-1.40)	(-1.46)	(-1.35)
R^2	0.05	0.69	1.04	1.47
U.S. Goverment Bond 10+ Years	-1.40	-0.86	-0.29	0.00
(t-stat)	(-0.59)	(-0.51)	(-0.20)	(0.00)
R^2	-0.29	-0.27	-0.40	-0.44

Table 10

Predicting Bond Returns (Cont.)

Panel B: Investment Grade Corporate Bond Indexes

	h=1	h=3	h=6	h=12
1993-2019				
Corporate Bond 0-1 Years	-0.13	-0.13 *	-0.11 *	-0.13 *
(t-stat)	(-1.28)	(-1.70)	(-1.72)	(-1.83)
R^2	-0.04	0.34	0.55	1.42
Corporate Bond 1-3 Years	-0.34	-0.35	-0.30	-0.24
(t-stat)	(-1.02)	(-1.35)	(-1.41)	(-1.10)
R^2	-0.06	0.30	0.50	0.60
Corporate Bond 3-5 Years	-0.16	-0.26	-0.21	-0.13
(t-stat)	(-0.27)	(-0.61)	(-0.61)	(-0.38)
R^2	-0.29	-0.19	-0.17	-0.23
Corporate Bond 5-7 Years	0.02	-0.27	-0.19	-0.00
(t-stat)	(0.03)	(-0.44)	(-0.38)	(-0.01)
R^2	-0.31	-0.24	-0.26	-0.32
Corporate Bond 7-10 Years	0.22	-0.23	-0.11	0.09
(t-stat)	(0.21)	(-0.30)	(-0.17)	(0.16)
R^2	-0.30	-0.28	-0.30	-0.31
Corporate Bond 10+ Years	0.62	-0.14	0.15	0.34
(t-stat)	(0.39)	(-0.12)	(0.14)	(0.44)
R^2	-0.27	-0.31	-0.30	-0.19
2000-2019				
Corporate Bond 0-1 Years	-0.26 *	-0.26 **	-0.25 **	-0.27 **
(t-stat)	(-1.81)	(-2.14)	(-2.31)	(-2.58)
R^2	0.42	1.63	2.85	5.29
Corporate Bond 1-3 Years	-0.54	-0.58 *	-0.58 *	-0.47
(t-stat)	(-1.30)	(-1.82)	(-1.91)	(-1.51)
R^2	0.13	1.07	2.20	2.59
Corporate Bond 3-5 Years	-0.45	-0.64	-0.69	-0.53
(t-stat)	(-0.64)	(-1.28)	(-1.53)	(-1.15)
R^2	-0.28	0.29	1.02	1.03
Corporate Bond 5-7 Years	-0.34	-0.75	-0.80	-0.51
(t-stat)	(-0.35)	(-1.09)	(-1.26)	(-0.79)
R^2	-0.38	0.08	0.53	0.21
Corporate Bond 7-10 Years	-0.22	-0.79	-0.77	-0.45
(t-stat)	(-0.17)	(-0.90)	(-0.99)	(-0.62)
R^2	-0.41	-0.04	0.25	-0.02
Corporate Bond 10+ Years	0.17	-0.61	-0.46	-0.18
(t-stat)	(0.09)	(-0.46)	(-0.38)	(-0.19)
R^2	-0.42	-0.32	-0.32	-0.41

Table 10
Predicting Bond Returns (Cont.)

Panel C: High Yield Corporate Bond Indexes

	h=1	h=3	h=6	h=12
	1993-2019			
Corporate Bond 0-1 Years	0.56	0.22	-0.04	-0.14
(t-stat)	(0.81)	(0.54)	(-0.12)	(-0.35)
R^2	-0.18	-0.25	-0.31	-0.25
Corporate Bond 1-3 Years	1.34	0.81	0.53	0.30
(t-stat)	(1.42)	(1.05)	(0.74)	(0.43)
R^2	0.10	0.02	-0.11	-0.20
Corporate Bond 3-5 Years	2.68 **	1.69 *	1.30	1.14
(t-stat)	(2.21)	(1.69)	(1.35)	(1.05)
R^2	1.02	0.70	0.57	0.86
Corporate Bond 5-7 Years	3.48 **	2.03 *	1.77 *	1.65
(t-stat)	(2.49)	(1.87)	(1.72)	(1.43)
R^2	1.42	0.87	1.02	1.74
Corporate Bond 7-10 Years	3.33 **	1.73	1.54	1.62
(t-stat)	(2.16)	(1.54)	(1.46)	(1.34)
R^2	0.91	0.46	0.68	1.66
Corporate Bond 10+ Years	2.73 *	1.50	1.03	0.88
(t-stat)	(1.70)	(1.19)	(0.83)	(0.70)
R^2	0.42	0.10	-0.04	0.01
	2000-2019			
Corporate Bond 0-1 Years	0.32	-0.11	-0.49	-0.65
(t-stat)	(0.30)	(-0.20)	(-1.02)	(-1.37)
R^2	-0.39	-0.41	-0.05	0.88
Corporate Bond 1-3 Years	1.19	0.29	-0.20	-0.54
(t-stat)	(0.87)	(0.29)	(-0.20)	(-0.55)
R^2	-0.17	-0.39	-0.41	-0.13
Corporate Bond 3-5 Years	2.98 *	1.60	0.91	0.41
(t-stat)	(1.89)	(1.37)	(0.77)	(0.30)
R^2	0.90	0.31	-0.08	-0.32
Corporate Bond 5-7 Years	3.95 **	1.89	1.37	0.94
(t-stat)	(2.12)	(1.41)	(1.01)	(0.58)
R^2	1.35	0.39	0.20	0.10
Corporate Bond 7-10 Years	4.48 **	2.22	1.73	1.55
(t-stat)	(2.23)	(1.64)	(1.23)	(0.88)
R^2	1.38	0.60	0.57	1.01
Corporate Bond 10+ Years	2.89	1.15	0.13	-0.20
(t-stat)	(1.39)	(0.78)	(0.08)	(-0.12)
R^2	0.25	-0.23	-0.43	-0.43

Table 11
Predicting Bond Returns: Out-Of-Sample R^2

This table reports the out-of-sample R^2 (R^2_{OS}) statistic (Campbell and Thompson 2008) in predicting the excess returns over the next h months on Datastream Treasury bond indexes (Panel A), S&P investment grade corporate bond (Panel B), and S&P high yield corporate bond (Panel C) using NYT *War*. 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 2003, and the whole sample is from January 1993 to October 2019. $*p < 0.1$; $**p < 0.05$; $***p < 0.01$ (based on the Clark and West (2007) MSFE-adjusted statistic).

	h=1	h=3	h=6	h=12
Panel A: Government Bond Indexes				
1-3 Years	0.21 **	0.18 *	3.38 **	7.31 **
3-5 Years	-1.39	-1.16	-0.76	1.23
5-7 Years	-1.57	-1.73	-1.90	-0.88
7-10 Years	-1.50	-1.87	-3.07	-2.15
10+ Years	-0.86	-1.73	-3.65	-2.08
Panel B: Investment Grade Corporate Bond Indexes				
0-1 Years	-0.27	0.29 *	0.46 *	0.65
1-3 Years	-0.84	-0.72	0.49	0.05
3-5 Years	-1.29	-1.28	-0.61	-1.54
5-7 Years	-1.19	-1.49	-0.94	-1.52
7-10 Years	-0.94	-1.43	-1.45	-1.40
10+ Years	-0.62	-1.31	-1.95	-1.27
Panel C: High Yield Corporate Bond Indexes				
0-1 Years	-0.07	-1.16	-0.79	-1.16
1-3 Years	0.02	-1.45	-1.54	-1.88
3-5 Years	0.30 *	-1.82	-3.06	-5.86
5-7 Years	0.44 **	-1.99	-3.36	-5.93
7-10 Years	0.58 **	-1.31	-1.64	-3.10
10+ Years	0.36 *	-1.04	-1.23	-1.16

Internet Appendix

War Discourse and Disaster Premia: 160 Years of Evidence from Stock and Bond Markets

Internet Appendix **2**

A Seeded Latent Dirichlet Distribution **2**

B Text Processing Steps **5**

C Additional Figures and Tables for Discourse Topics from the *NYT* **7**

C.1 *War* versus Conditional Volatility and Skewness 7

C.2 *War* as a Proxy for Time-Varying Risk Aversion 8

C.3 *War* versus Actual Events 11

C.4 Subperiod Predicting Power 12

C.5 Predicting Returns on Characteristic Portfolios 14

D Topic Weights Constructed by Raw Counts of Seed Words from the *NYT* **26**

E Discourse Topics from the *WSJ* **33**

A Seeded Latent Dirichlet Distribution

In this appendix, we provide more details on the seeded latent Dirichlet distribution model. This paper uses a stochastic topic model to extract latent topic weights from news articles. Topic models are developed based on the core idea that documents are mixtures of topics, where each topic has a probability distribution over words (Blei 2012; Steyvers and Griffiths 2007). Under topic models, we assume that text documents derive from a stochastic generative process. The creation of a new document starts with a document-specific distribution over topics (the document-topic distribution). Each word in the document is chosen first by picking a topic randomly from the document-topic distribution, and then drawing a word from the topic-word distribution for that topic. To model this, every possible word needs to be assigned to a topic.

In this setup, the document-topic distribution for each document and topic-word distribution for each topic (the same across documents) are unobserved parameters that are estimated from the observable word frequencies in the document collection. In other words, we can use standard statistical techniques to estimate the generative process, inferring the topics responsible for generating a collection of documents (Steyvers and Griffiths 2007).

The most widely used topic model is latent Dirichlet allocation (LDA) as introduced by Blei, Ng, and Jordan (2003) and further developed by Griffiths and Steyvers (2004). Under LDA, a document d is generated under the following hierarchical process:

- The word weight vector ω_k of topic k is the vector of probabilities of each word value for the topic k. The prior for these weights is assumed to have a Dirichlet distribution governed by parameter β: $\omega_k \sim \text{Dirichlet}(\beta)$.[1]
- The topic weight in a document d, denoted τ_d, is a vector of topic probabilities, i.e., probabilities that any given word location in the document is about any given topic. The topic weight vector of document d follows a prior Dirichlet distribution governed by parameter α, the same for all documents: $\tau_d \sim \text{Dirichlet}(\alpha)$, the same for all documents.[2]

[1] To illustrate, suppose that topic k has three words: $word_1$, $word_2$, and $word_3$ with respective weights $\omega_k = [w_1, w_2, w_3]$ with $w_1 + w_2 + w_3 = 1$. The model assumes that this ω_k vector follows a Dirichlet distribution.

[2] Similarly, assume document d has four topics $topic_1$, $topic_2$, $topic_3$, $topic_4$ with the weights given to these topics captured by $\tau_d = [\theta_1, \theta_2, \theta_3, \theta_4]$ with $\theta_1 + \theta_2 + \theta_3 + \theta_4 = 1$. The model assumes that this τ_d vector follows a Dirichlet distribution.

- We use v to indicate a word location in a given document, and w to indicate a word value (such as "the" or "cat"). For each word location v in document d, we
  - randomly select a topic from the document-topic distribution:
    $z_{dv} \sim \text{Multinomial}(\tau_d)$ (a distribution which does not depend on v), and then
  - randomly select from a word from that topic:
    $w \sim \text{Multinomial}(\omega_{z_{dv}})$.

In other words, it is the multinomial distribution of word values for the realized topic z_{dv}.

In this setup, the topic-word distribution ω_k and document-topic distribution τ_d are latent parameters that we want to estimate. Estimating these involves a backward inference based on observed word frequencies across documents. The parameters α and β are hyperparameters of the prior distribution whose values are taken from the Latent Dirichlet Distribution topic modelling literature.

The document-topic distribution τ_d is of utmost interest because it summarizes the attention allocated to each topic in each news article. To estimate these parameters using a Bayesian method, Griffiths and Steyvers (2004) specifies that ω_k and τ_d follow two Dirichlet distributions (these two are referred to as the "prior" distribution in Bayesian statistic). From these specifications, we can derive the distribution of the topic assignment z_{dv} conditioned on observed word frequencies (this conditional distribution is referred to as the "posterior" distribution). We then use Gibbs sampling to simulate this posterior distribution and estimate the two hidden model parameters.[3]

Users of the traditional unsupervised LDA developed by Blei, Ng, and Jordan (2003) and Griffiths and Steyvers (2004) only need to prespecify the number of topics K and let the model cluster words into these topics based on word frequencies in a completely unsupervised manner. Specifically, the LDA model is more likely to assign a word w to a topic k in a document d if w has been assigned to k across many different documents and k has been used multiple times in d (Steyvers and Griffiths 2007). The model automatically extracts underlying topics, so users of LDA have no control over topic assignments.

[3]Gibbs sampling is a sampling technique to simulate a high-dimensional distribution by sampling from lower-dimensional subsets of variables where each subset is conditioned on the value of all others. See Griffiths and Steyvers (2004) for details on the implementation of Gibbs sampling in LDA.

Since we are interested in uncovering some specific topics, we employ a recent extension of LDA called seeded LDA (sLDA) developed by Lu et al. (2011). sLDA allows users to regulate topic contents using domain knowledge by injecting seed words (prior knowledge) into the model. Precisely, under sLDA, we specify the topic-word distribution as follows:

$$\omega_k \sim \text{Dirichlet}\,(\beta + C_w)_{w \in V}\,, \tag{A.1}$$

where V is the corpus or text collection, $C_w > 0$ when w is a seed word in topic k and $C_w = 0$ when w is not a seed word. The higher is C_w, the stronger the tilt toward word w appearing in any given topic. Intuitively, sLDA gives preference to seed words w in topic k in the form of pseudo count C_w and clusters words into topics based on their co-occurrences with the seed words. When a seed word is not present in a text collection, it does not enter the sLDA model and has no impact on the estimation process.

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. Following standard practice, we set $\alpha = 50/K$ where K is the number of topics, $\beta = 0.1$, and $C_w = 0.01$ times the number of terms in the corpus.

B Text Processing Steps

Before carrying out text cleaning, we first remove articles with limited contents, i.e., articles containing mostly numbers, names, lists, programs, etc.

We manually check and infer title patterns that indicate limited content. About 1.4 million articles have limited content out of the total 14.7 million articles as shown in Table C.1. List of exclusion patterns are available from authors on requests.

Next, we conduct the following text cleaning steps:

[1] Remove articles with fewer than 100 content words. We consider content words as those outside 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 documents 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 essential to retain punctuations before tokenization. Keeping punctuations and stop words before creating n-grams ensures that our ngrams are present in the corpus. An alternative approach is to tokenize texts after removing punctuations and stop words. However, this approach results in n-grams that do not appear in the documents, thus distorting the original thematic contents of the document.

[5] Remove unigrams of fewer than three letters or being a stop word and bigrams containing stop words. We also remove trigrams containing stopwords unless the stop word is a

preposition in the middle position. For example, under within-punctuation boundary tokenization, the sentence "Under current favorable conditions, the 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, revenue_of_firm] where all stop words and words of less than three characters have been removed. Meaningless ngrams that have stop words on the boundaries such as under_current_favorable (which does not add any additional meaning to current_favorable) have been removed while revenue_of_firm is retained. We also experiment with keeping stop words with future meaning, such as [will, might, could, should, possible, likely, forward, future, pending, etc.], and obtain similar results.

[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 an ngram or token, and each entry is the count of the token in that document. We put all ngrams into one document-frequency matrix. 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 during each estimation window.

C Additional Figures and Tables for Discourse Topics from the *NYT*

This appendix reports additional tables and figures for the discourse topics constructed from the *NYT*. Figure C.1 plots the monthly count and monthly article length of our *NYT* data set. Figure C.2 shows the word clouds for the remaining eight topics, and Figure C.3 shows their time series.

Table C.1 reports the number of *NYT* articles left after each screening step. Table C.2 reports the correlation matrix of the topics and the PLS index.

C.1 *War* versus Conditional Volatility and Skewness

Our *War* index could possibly capture conditional market return, volatility, and skewness that have market return predictability. To investigate this possibility, we re-run our predictive regression and control for variables:

$$R^e_{t+1} = \alpha + \beta War_t + \gamma z_t + \epsilon_{t+1}, \tag{C.1}$$

where z_t is either the current market excess return, conditional volatility, or conditional skewness. We construct the monthly conditional market volatility, $\hat{\sigma}_t$, from daily returns as follows:

$$\hat{\sigma}_t = \sqrt{\frac{1}{n_t - 1}\sum_{\tau=1}^{n_t}(R_\tau - \bar{R}_t)^2}, \tag{C.2}$$

where n_t is the number of trading days in month t, R_τ is the daily return and $\bar{R}_t$ is the average daily returns in month t. Similarly, we construct the monthly conditional skewness as follows:

$$sk_t = \frac{n_t}{(n_t-1)(n_t-2)}\sum_{\tau=1}^{n_t}\Big(\frac{R_\tau - \bar{R}_t}{\hat{\sigma}_t}\Big)^3, \tag{C.3}$$

where following standard practice we scale the raw central third moment by the standard deviation. Because daily data on the S&P 500 index becomes available in January 1928, our sample is from January 1928 to October 2019.

We report the results in Table C.3. Panel A reports the results for the whole sample from 1928 to 2019. We find that the predictability of *War* is not affected by any of the return moments. When we control for all three moments in the last column, the predictive power of *War* is still intact. We obtain similar results over two subsamples: 1950-2019 and 2000-2019. This result confirms that the predictability of *War* does not come from other return moments.

C.2 *War* as a Proxy for Time-Varying Risk Aversion

In the main text, we show that *War* captures rare disaster probability as *War* is a positive market predictor, and innovations in *War* command a negative risk premium. These results are consistent with the predictions of the rare disaster risk model. In this section, we further show that *War* proxies for time-varying risk aversion, lending empirical support to the ICAPM. We first briefly discuss the ICAPM framework and then present the empirical results.

Before hypothesizing that *War* captures time-varying risk aversion, we first briefly introduce the Merton (1973)'s 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, $\sigma_{MF,t}$:

$$\mathbb{E}_t[R_{M,t+1} - R_{f,t+1}] = \left[\frac{-J_{WW}W}{J_W}\right]\sigma^2_{M,t} + \left[\frac{-J_{WF}}{J_W}\right]\sigma_{MF,t}, \tag{C.4}$$

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 (C.4) 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 (C.4) 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 where 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 (C.4):

$$\mathbb{E}_t[R_{M,t+1} - R_{f,t+1}] = \lambda_0 + \lambda_1 \times \sigma^2_{M,t}, \tag{C.5}$$

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 (C.5) 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 return-volatility trade-off:

$$R_{M,t+1} - R_{f,t+1} = \lambda_0 + \lambda_1 \times \sigma^2_{M,t} + \epsilon_{t+1}, \tag{C.6}$$

where ϵ_{t+1} has a mean of zero with conditional variance $\sigma^2_{M,t}$. Empirical tests of Equation (C.6) on the U.S. stock market return have 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 (C.5) and Equation (C.6) assumes that the coefficient of relative risk aversion, λ_1, is time-invariant. However, we have no compelling reason to believe this assumption would hold in practice. Indeed, relative risk aversion is modeled as time-varying in several asset pricing models, such as the external habit model by Campbell and Cochrane (1999). 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:

$$R_{M,t+1} - R_{f,t+1} = \lambda_0 + (\lambda_1 + \lambda_2 \times x_t) \times \sigma^2_{M,t} + \epsilon_{t+1}. \tag{C.7}$$

We hypothesize that *War* proxies for time-varying relative risk aversion, and, thus, we replace the state variable, x_t, with *War* in Equation (C.7). Hence, $\lambda_t = \lambda_1 + \lambda_2 \times War_t$. If this hypothesis holds with real-world data, then we expect (1) the adjusted R^2 of Equation (C.7) to be higher than that of (C.6), as the former is a more proper representation of the risk-return trade-off, and (2) the coefficient λ_2 in Equation (C.7) to be significantly positive as risk aversion is expected to rise when War 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):

$$
\begin{aligned}
\text{GARCH}(1,1): \sigma^2_{M,t} &= \delta_0 + \delta_1\epsilon^2_t + \delta_2\sigma^2_{M,t-1} \\
\text{IGARCH}(1,1): \quad \sigma^2_{M,t} &= \delta_0 + \delta_1\epsilon^2_t + (1-\delta_1)\sigma^2_{M,t-1} \\
\text{TGARCH}(1,1): \quad \sigma^2_{M,t} &= \delta_0 + \delta_1\epsilon^2_t + \delta_3 D_t\epsilon^2_t + \delta_2\sigma^2_{M,t-1} \\
\text{EGARCH}(1,1): \quad ln(\sigma^2_{M,t}) &= \delta_0 + \delta_1\left(\frac{|\epsilon_t|}{\sigma_{M,t}}\right) - \delta_3\left(\frac{\epsilon_t}{\sigma_{M,t}}\right) + \delta_2 ln(\sigma^2_{M,t-1}),
\end{aligned}
\tag{C.8}
$$

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

Panel A of Table C.4 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 1% level. Hence, we observe the positive risk-return trade-off with a large sample size. 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 War 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 2.05, significant at the 1% level, and the adjusted R^2 jumps from -0.38% to 0.27%, 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, and the coefficient λ_2 is significant at the 10% level. In the second subsample, R^2 jumps more than eight times, and λ_2 is significant at the 5% level under the time-varying RRA model.

Panels B, C, and D of Table C.4 report the results with different specifications for the volatility equation. We obtain consistent results across both the models and sample periods, except for EGARCH in the whole sample, confirming that *War* captures risk aversion, enhancing the risk-return relationship.

C.3 *War* versus Actual Events

As *War* is constructed to be the attention paid to wars and tensions, it is interesting to examine whether *War* has the predictive power beyond the actual rare disasters. To answer this question, we first create indicators for these events reported by GFD:

- Recessions: from NBER;
- Bank failures: if the event is tagged as bank failure, War, 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.

Figure C.4 plots these events over the past 150 years.

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

$$R^e_{t+1} = \alpha + \beta \times War_t + \gamma^j \times D^j_t + \epsilon_{t+1}, \tag{C.9}$$

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

Panel A of Table C.5 reports the results for the whole sample. Across all events, *War* remains significant as a return predictor. Among the events, only Recessions and Epidemic yield significant prediction coefficients at the 5% and 10% levels, respectively. The prediction slope on Recessions is negative and is thus inconsistent with a risk-based explanation. The results indicate that the actual events themselves, except Recessions, have limited predictive power and therefore cannot be a cause of fluctuations in RRA. This evidence rules out the possibility that *War* only reflects RRA changes triggered by real-world stressful events.

Panel B reports the results in the first half of the sample from 1871 to 1949. During this period, *War* remains significant against Bank Failures, Disasters, and Epidemic. During the second half of the sample, *War* remains significant at least 5% 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 *War* from news articles is simply a manifestation of actual events. Indeed, we find that most of the events have no predictive power. Thus, it is undoubtedly the narrative aspects of the events that matter for the stock market.

C.4 Subperiod Predicting Power

This subsection investigates the predictive power of discourse topics 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 (Huang et al., 2015; Jiang et al., 2019; Stambaugh, Yu, and Yuan, 2012). Huang, Li, and Wang (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, Li, and Wang, 2020). While these observations lean toward the behavioral channel, the predictability of our topics is more risk-based, so whether we can observe similar subsample concentrations in predictability remains unclear.

To examine the above question, we follow Rapach, Strauss, and Zhou (2010) and Huang et al.

(2015), among others. We compute the subsample R^2 as follows:

$$R_c^2 = 1 - \frac{\sum_{t=1}^{T} I_t^c \left(\hat{\epsilon}_t\right)^2}{\sum_{t=1}^{T} I_t^c \left(R_t^e - \bar{R}^e\right)^2}, \quad c = exp, \;\; rec, \;\; high, \;\; low, \tag{C.10}$$

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 the National Bureau of Economic Research (NBER) business cycles. For sentiment periods, we follow Stambaugh, Yu, and Yuan (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 (C.10), we compute the out-of-sample R_{OS}^2 for each period. Similar to the previous out-of-sample analyses, we use the expanding estimation window, and the evaluation period began in January 1891.

Panel A of Table C.6 reports the results with the in-sample R^2. Accordingly, *War* and the PLS index yield higher R^2's during recessions (0.91% in recessions vs. 0.06% in expansions for *War*, and 1.80% in recessions vs. 0.58% in expansions for PLS). These results are consistent with the observation of concentrated predictive power during recessions documented in the literature. However, the out-of-sample R^2 with an expanding window in Panel B suggests both *War* and the PLS index have stronger predictive power in expansions (0.69% in expansions vs. -0.46% in recessions for *War*, and 0.10% in expansions vs. -0.31% in recessions for PLS). In sum, whether topics have stronger prediction power in recessions remains inconclusive.

We consistently find that discourse topics can better predict the market during low sentiment periods for both in-sample and out-of-sample analyses. For example, the in-sample R^2 for *War* is 0.50% during low sentiment periods versus 0.01% during high sentiment periods, while the figure for PLS is 1.90% versus -0.16%, respectively. For out-of-sample prediction, *War* yields an R_{OS}^2 of 0.47% during low sentiment months versus -0.02% during high sentiment months, while the numbers for PLS are 0.85% and -0.66%. While this result is contradictory to the sentiment literature, it is intuitive. When people are in a bad mood, they are more receptive to stressful news.

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

C.5 Predicting Returns on Characteristic Portfolios

In the main text, we document that *War* and the discourse topic index predict market returns. In this appendix, we investigate whether the return predictability of topics holds at the individual portfolio level. 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 topics over the risk premium on the characteristics portfolios, we run the following predictive regression:

$$R^e_{i,t+1} = \alpha_i + \beta_i x_t + \epsilon_{i,t+1}, \quad i = 1, \ldots, 40 \tag{C.11}$$

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

Panel A of Table C.7 reports results with 10 industry portfolios. Both *War* and the PLS index yield positive slope coefficients across industries, although most of the prediction coefficients on *War* are insignificant. 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 C.7 reports results with the size, BM, and momentum portfolios. Both *War* and the PLS index yield positive slopes for these portfolios, but the prediction coefficients on *War* are not as strong as those on the PLS index. The slopes on the ten-size portfolios increase monotonically from the large to small portfolios for both *War* and the PLS index. The topics also better predict value (high BM) and past loser stocks. Thus, returns on small, distressed (high BM) and, recently, underperforming stocks are more sensitive to War.

Figure C.1. ***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. The sample period is from January 1871 to October 2019. Articles with limited content have been removed.

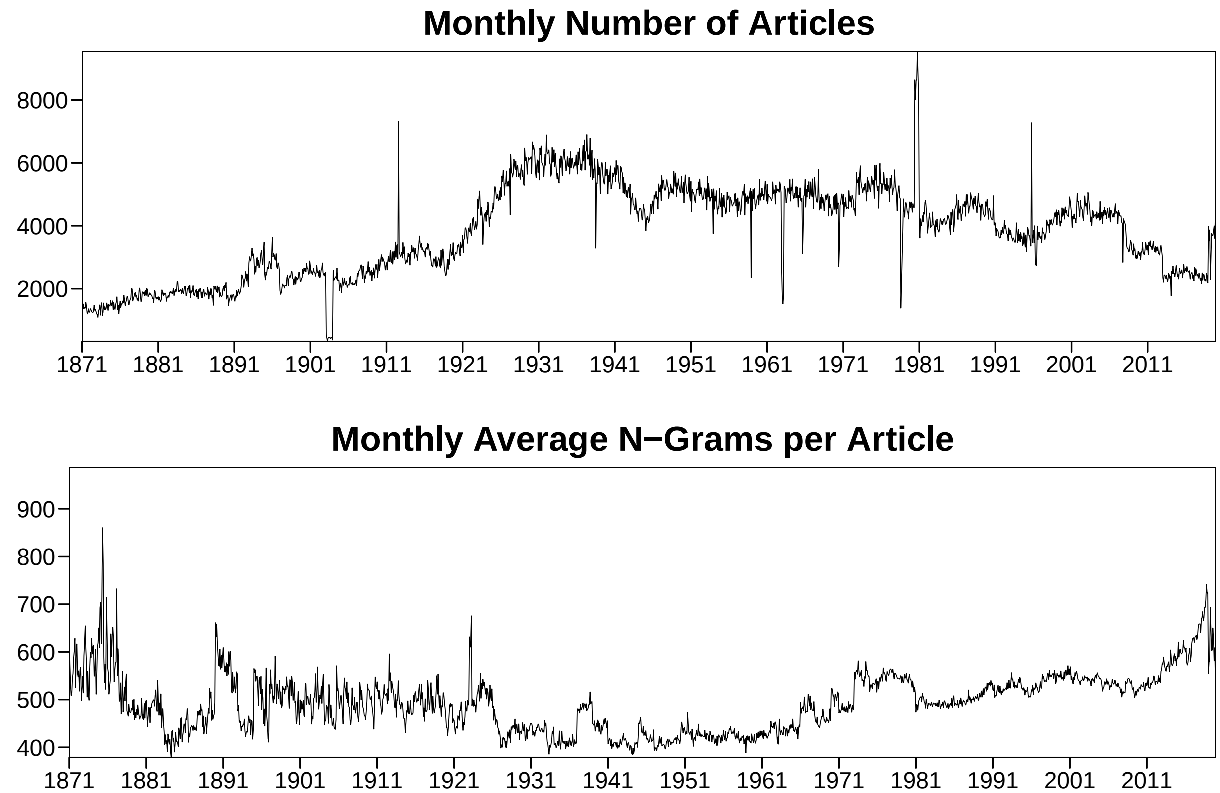
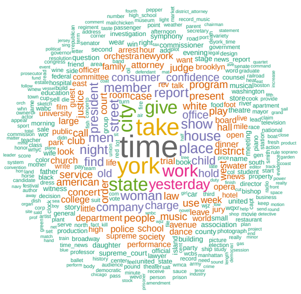


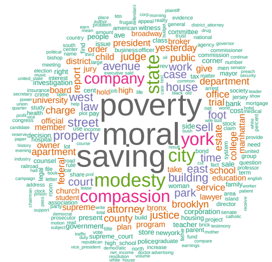
Figure C.2. Discourse Topic Contents

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

Confidence

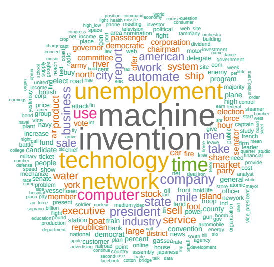


Savings

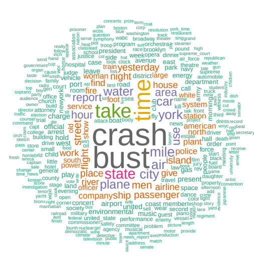


Conspicuous Consumption

Technology Replacing Jobs

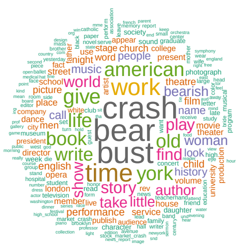


Real Estate Crashes



Stock Market Bubbles

Stock Market Crashes



Wage and Labor Unions

Figure C.3. Time Series of Topics 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.

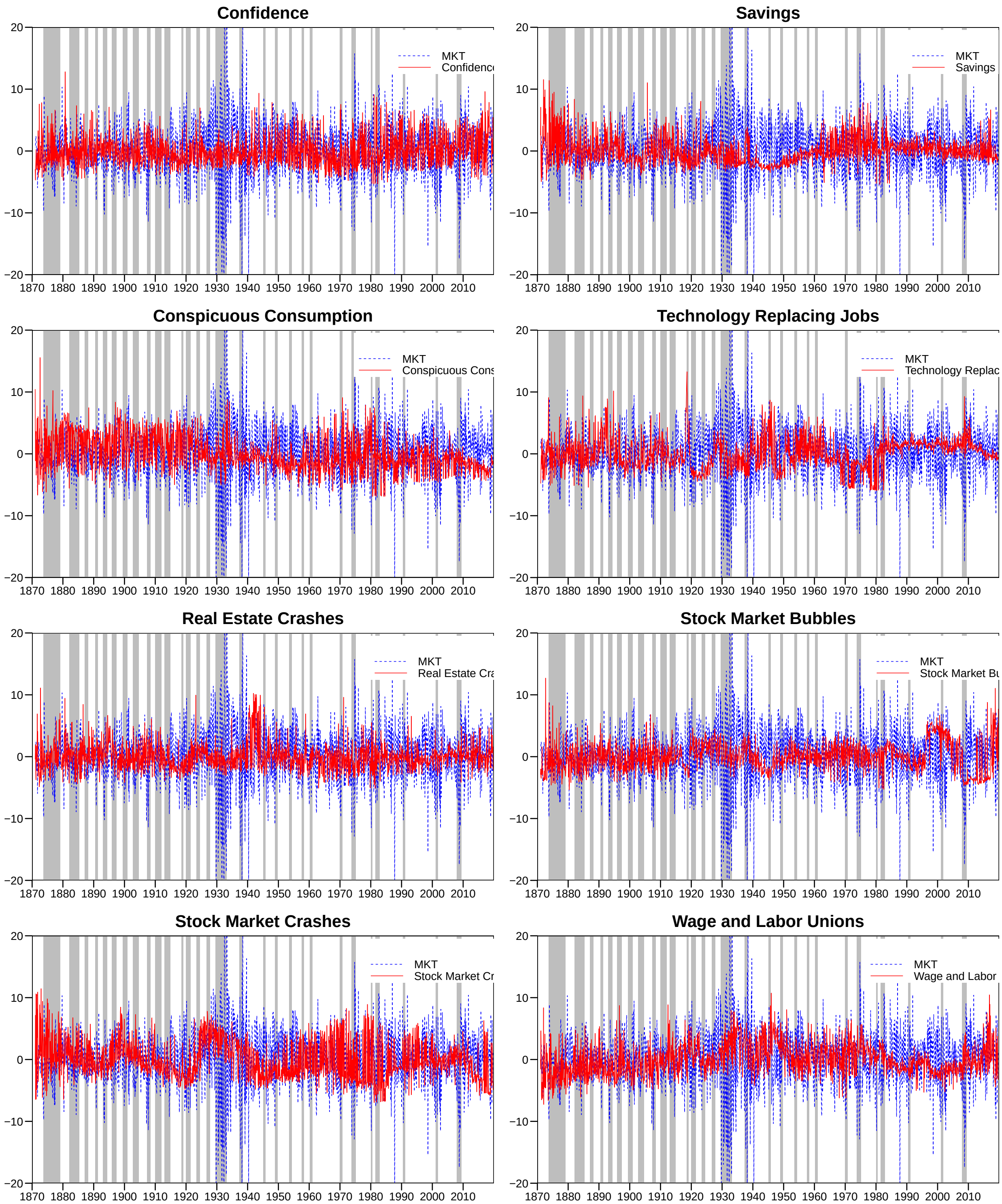


Figure C.4. Historical Events

This figure plots the time series of monthly historical U.S. events. The horizontal gray bar indicates at least one 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.

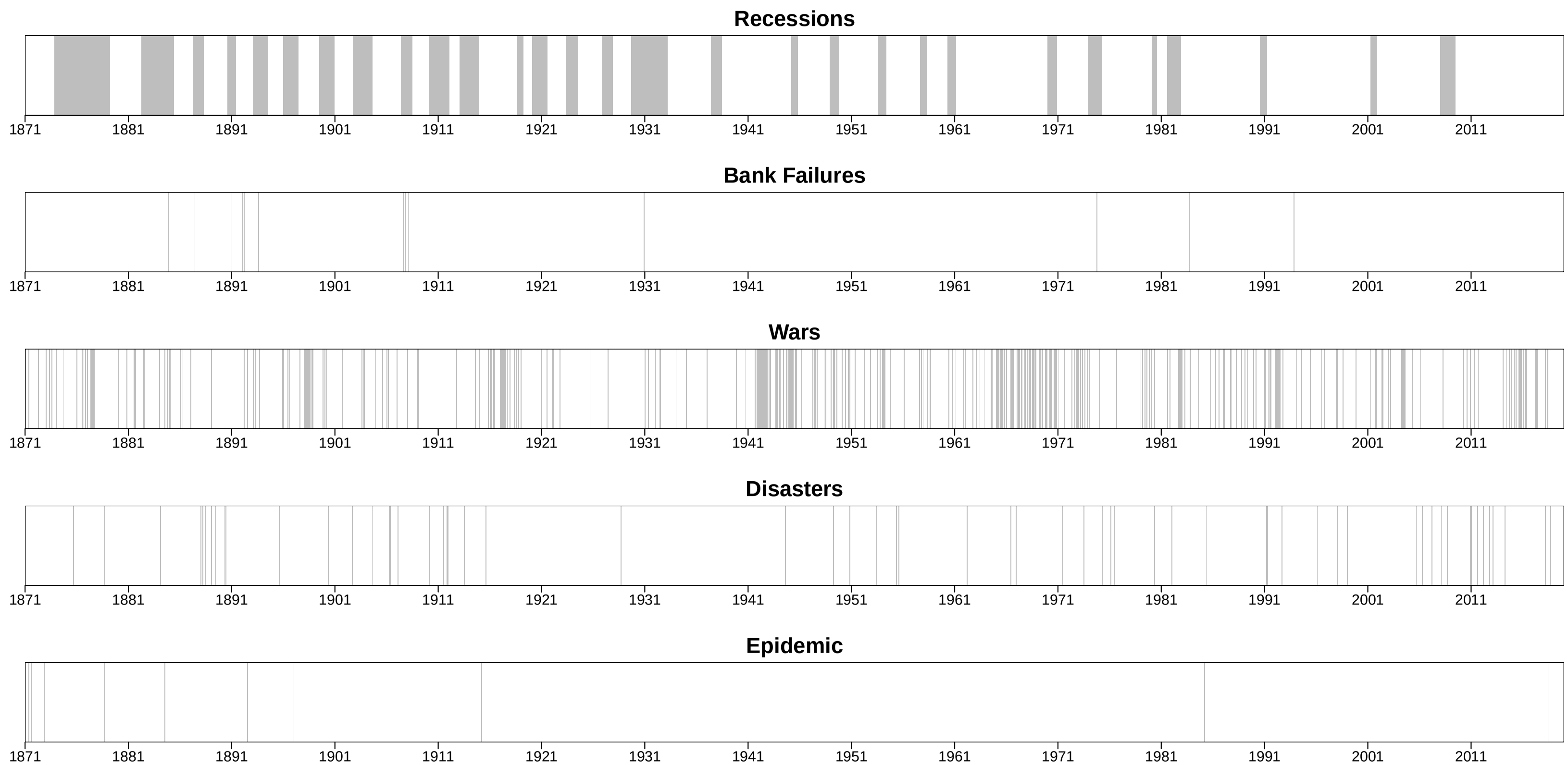


Table C.1
Data Screening

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 C.2
Topic Correlations

This table presents the pairwise correlations among 14 monthly topic weights constructed according to the sLDA model described in Section 2 and the PLS index. All numbers (except sample size) are expressed as percentages. The sample period is from January 1871 to October 2019.

	War	Pandemic	Panic	Confidence	Saving	Consumption	Money	Tech	Real estate boom	Real estate crash	Stock bubble	Stock crash	Boycott	Wage	PLS
War		-6.5	-17.1	-7.9	-22.2	-26.2	-10.0	-1.5	-1.7	-10.1	-10.1	-19.9	-32.3	22.7	82.1
Pandemic	-6.5		-6.5	-10.7	-2.2	-8.8	-2.3	-4.5	-7.3	-1.2	-3.7	-13.9	-2.2	-9.0	-28.5
Panic	-17.1	-6.5		-6.2	-14.9	5.3	-3.5	-8.2	-6.1	-3.0	-15.3	-14.1	-3.4	0.7	15.7
Confidence	-7.9	-10.7	-6.2		-3.3	-10.0	-8.0	6.6	-8.5	-1.6	-5.9	-20.0	-0.1	-11.5	-0.5
Saving	-22.2	-2.2	-14.9	-3.3		-5.8	-4.9	-6.8	-0.9	-7.3	-0.8	3.1	16.0	-22.3	-34.1
Consumption	-26.2	-8.8	5.3	-10.0	-5.8		10.4	-14.4	-17.9	-4.3	-7.0	-1.2	2.7	-13.5	-4.6
Money	-10.0	-2.3	-3.5	-8.0	-4.9	10.4		-14.6	-9.3	-12.4	-10.8	1.8	-13.5	5.9	-15.3
Tech	-1.5	-4.5	-8.2	6.6	-6.8	-14.4	-14.6		-4.1	-2.8	-8.6	-7.0	-14.2	-22.4	-8.1
Real estate boom	-1.7	-7.3	-6.1	-8.5	-0.9	-17.9	-9.3	-4.1		-1.8	-0.9	-17.4	-2.6	-6.7	-32.7
Real estate crash	-10.1	-1.2	-3.0	-1.6	-7.3	-4.3	-12.4	-2.8	-1.8		-7.7	-6.0	-10.0	-6.2	-1.1
Stock bubble	-10.1	-3.7	-15.3	-5.9	-0.8	-7.0	-10.8	-8.6	-0.9	-7.7		-0.0	-0.8	-13.8	-14.7
Stock crash	-19.9	-13.9	-14.1	-20.0	3.1	-1.2	1.8	-7.0	-17.4	-6.0	-0.0		-0.3	-5.6	-2.3
Boycott	-32.3	-2.2	-3.4	-0.1	16.0	2.7	-13.5	-14.2	-2.6	-10.0	-0.8	-0.3		-24.4	-41.7
Wage	22.7	-9.0	0.7	-11.5	-22.3	-13.5	5.9	-22.4	-6.7	-6.2	-13.8	-5.6	-24.4		37.3
PLS	82.1	-28.5	15.7	-0.5	-34.1	-4.6	-15.3	-8.1	-32.7	-1.1	-14.7	-2.3	-41.7	37.3	

Table C.3
Predicting One-Month Market Returns: *War* versus Return Moments

This table presents the results of the following predictive regression:

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

where R^e_{t+1} is the excess market return over the next month, War_t is the *NYT War* index, z_t is one of the current excess market return (MKT), conditional volatility (VOL), or conditional skewness (SK), and β 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. *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$. The whole sample is from January 1928 to October 2019.

	Panel A: 1928-2019			
War	3.40 **	3.71 **	3.48 **	3.61 **
	(2.13)	(2.27)	(2.16)	(2.27)
MKT	5.28			6.22
	(1.39)			(1.46)
VOL		0.68		2.31
		(0.14)		(0.44)
SK			-2.33	-3.27
			(-0.89)	(-1.37)
R^2	0.78	0.13	0.25	0.93
	Panel B: 1950-2019			
War	3.97 **	3.71 **	4.00 **	3.63 **
	(2.54)	(2.33)	(2.53)	(2.30)
MKT	1.44			1.10
	(0.66)			(0.53)
VOL		-1.92		-1.62
		(-0.67)		(-0.57)
SK			-1.29	-1.45
			(-0.74)	(-0.81)
R^2	0.52	0.58	0.50	0.46
	Panel C: 2000-2019			
War	9.49 ***	8.81 ***	9.75 ***	8.71 ***
	(3.31)	(2.92)	(3.39)	(2.86)
MKT	2.27			0.36
	(0.55)			(0.09)
VOL		-5.10		-4.94
		(-0.95)		(-0.86)
SK			-1.01	-0.99
			(-0.38)	(-0.37)
R^2	3.18	3.96	3.02	3.18

Table C.4
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):

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

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 War, is in bold. The whole sample is from 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	-0.85	0.00	0.37	0.00	-0.72	0.00	1.09	0.00	-0.83
λ_1	2.17	2.62	0.22	0.21	2.20	2.42	0.81	0.68	2.58	1.32	-0.78	-0.33
λ_2			**2.05**	**3.20**			**1.56**	**1.89**			**3.10**	**2.46**
δ_0	0.00	3.56	0.00	3.64	0.00	2.67	0.00	2.73	0.00	2.47	0.00	2.45
δ_1	0.14	6.42	0.14	6.60	0.16	4.92	0.17	5.08	0.12	3.75	0.12	3.72
δ_2	0.82	31.85	0.82	32.67	0.81	20.95	0.81	21.66	0.83	24.50	0.83	23.73
Adj. R^2(%)	-0.38		0.27		-0.73		-0.30		0.10		0.84	
Panel B: IGARCH												
λ_0	0.00	2.17	0.00	-0.73	0.00	0.68	0.00	-0.65	0.00	1.95	0.00	-0.64
λ_1	1.69	2.84	-0.03	-0.03	1.81	2.41	0.56	0.56	1.88	1.80	-1.18	-0.77
λ_2			**1.95**	**4.08**			**1.50**	**2.48**			**2.79**	**2.57**
δ_0	0.00	3.60	0.00	3.63	0.00	2.57	0.00	2.62	0.00	2.93	0.00	2.88
δ_1	0.18	6.65	0.18	6.76	0.20	4.65	0.20	4.85	0.16	5.50	0.16	5.36
δ_2	0.82		0.82		0.80		0.80		0.84		0.84	
Adj. R^2(%)	-0.32		0.29		-0.61		-0.21		0.11		0.76	
Panel C: TGARCH												
λ_0	0.00	2.11	0.00	-1.15	0.00	0.43	0.00	-2.05	0.01	4.19	0.00	-0.19
λ_1	2.33	10.18	0.07	0.09	2.09	2.02	1.06	1.32	0.31	0.33	-2.09	-0.96
λ_2			**2.17**	**9.34**			**1.77**	**6.80**			**2.55**	**4.82**
δ_0	0.00	6.48	0.00	4.93	0.00	4.01	0.00	2.81	0.01	1.33	0.01	1.65
δ_1	0.14	11.79	0.14	13.85	0.14	8.72	0.15	10.77	0.12	4.52	0.12	4.95
δ_2	0.84	805.33	0.84	75.72	0.85	616.62	0.84	43.90	0.76	7.04	0.76	9.01
δ_3	0.26	3.38	0.28	3.24	0.20	2.29	0.21	2.22	0.76	1.71	0.68	2.03
Adj. R^2(%)	-0.26		0.13		-0.35		-1.07		-0.11		0.42	
Panel D: EGARCH												
λ_0	0.00	1.17	0.00	-0.12	0.00	-0.95	0.00	-1.18	0.01	1.62	0.00	-2.02
λ_1	2.88	43.43	0.23	0.03	2.81	16.15	0.77	4.50	0.89	0.44	-1.92	-1.98
λ_2			**2.03**	**1.15**			**1.47**	**3.47**			**2.69**	**5.58**
δ_0	-0.28	-43.00	-0.28	-0.97	-0.21	-23.38	-0.19	-3.45	-0.81	-2.77	-0.80	-8.26
δ_1	-0.06	-3.75	-0.06	-0.72	-0.05	-2.38	-0.05	-2.40	-0.14	-2.59	-0.13	-4.25
δ_2	0.96	65420.25	0.95	20.30	0.97	4136.73	0.97	112.70	0.88	19.74	0.88	59.15
δ_3	0.26	8.27	0.26	2.35	0.26	5.47	0.25	5.69	0.22	5.46	0.22	5.71
Adj. R^2(%)	-0.56		0.20		-1.17		-0.17		-0.13		0.50	

Table C.5
War versus Real Events

This table presents the results of the following predictive regression:

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

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 War 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						
War	3.02 **	3.82 ***	3.58 ***	3.75 ***	3.84 ***	3.55 ***
	(2.52)	(3.37)	(3.06)	(3.32)	(3.40)	(3.14)
Event	-8.33 **	3.36	2.90	-7.06	18.32 *	-8.81 ***
	(-2.08)	(0.30)	(0.72)	(-1.10)	(1.66)	(-3.07)
R^2(%)	0.74	0.33	0.37	0.38	0.39	0.92
Panel B: 1871-1949						
War	2.54	3.62 **	2.68	3.42 **	3.54 **	3.35 *
	(1.37)	(2.09)	(1.45)	(1.98)	(2.05)	(1.93)
Event	-11.24 **	12.69	9.36	-8.11	20.43	-11.65 ***
	(-2.29)	(1.07)	(1.32)	(-1.15)	(1.57)	(-2.85)
R^2(%)	0.85	0.14	0.38	0.14	0.19	0.95
Panel C: 1950-2019						
War	4.12 **	4.02 **	4.16 ***	4.01 **	4.07 **	4.26 ***
	(2.58)	(2.53)	(2.62)	(2.53)	(2.57)	(2.66)
Event	-4.55	-32.15 *	-2.44	-6.22	9.74	-6.19
	(-0.64)	(-1.80)	(-0.56)	(-0.64)	(0.51)	(-1.57)
R^2(%)	0.53	0.58	0.48	0.50	0.44	0.79

Table C.6
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 the 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 entire 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					
War	0.39	0.06	0.91	0.01	0.50
PLS	1.07	0.58	1.80	-0.16	1.90
Panel B: Out of Sample					
War	0.17	0.69	-0.46	-0.02	0.47
PLS	-0.08	0.10	-0.31	-0.66	0.85

Table C.7
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 *War* 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. *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

	War(%)	t-stat	R^2 (%)	PLS (%)	t-stat	R^2 (%)
Panel A: Industry Portfolios						
Nondurable	2.53 *	(1.80)	0.12	4.07 ***	(2.64)	0.46
Durable	3.29	(1.42)	0.04	6.96 ***	(2.82)	0.48
Manufacture	2.16	(1.21)	-0.01	5.25 ***	(2.58)	0.40
Energy	1.69	(0.89)	-0.04	4.75 **	(2.30)	0.33
Technology	3.62	(1.57)	0.09	6.54 ***	(2.61)	0.48
Telecom	0.98	(0.70)	-0.06	2.43	(1.57)	0.11
Shop	3.18 *	(1.77)	0.12	6.01 ***	(3.34)	0.66
Health	2.17	(1.24)	0.02	4.95 ***	(2.65)	0.46
Utility	2.23	(1.35)	0.02	4.19 **	(2.42)	0.32
Other	3.75 **	(1.98)	0.15	6.38 ***	(2.92)	0.60
Panel B: Size Portfolios						
Small	6.17 **	(1.97)	0.18	10.67 ***	(2.83)	0.72
2	4.99 *	(1.88)	0.14	8.76 ***	(2.85)	0.62
3	5.32 **	(2.22)	0.23	8.39 ***	(3.02)	0.69
4	4.39 **	(1.97)	0.16	7.93 ***	(3.13)	0.71
5	4.20 **	(1.98)	0.16	7.06 ***	(2.99)	0.62
6	3.63 *	(1.78)	0.11	6.90 ***	(3.03)	0.64
7	3.91 **	(1.98)	0.17	6.71 ***	(3.11)	0.68
8	3.67 **	(2.01)	0.16	6.33 ***	(3.08)	0.66
9	3.13 *	(1.83)	0.11	5.93 ***	(3.10)	0.65
Large	2.54 *	(1.69)	0.09	4.90 ***	(2.93)	0.57
Panel C: Book-to-market Portfolios						
Growth	2.33	(1.37)	0.03	5.35 ***	(2.80)	0.53
2	2.49	(1.50)	0.06	5.11 ***	(2.88)	0.56
3	2.36	(1.48)	0.04	4.75 ***	(2.76)	0.46
4	2.01	(1.23)	-0.01	4.83 **	(2.50)	0.38
5	3.15 *	(1.90)	0.13	5.88 ***	(3.15)	0.67
6	2.30	(1.34)	0.01	5.15 **	(2.54)	0.42
7	3.47 *	(1.86)	0.11	5.89 ***	(2.79)	0.50
8	3.78 *	(1.92)	0.13	7.02 ***	(3.16)	0.68
9	4.54 **	(2.04)	0.16	7.77 ***	(3.09)	0.63
Value	5.77 **	(2.08)	0.19	9.08 ***	(2.88)	0.60
Panel D: Momentum Portfolios						
Losers	7.10 **	(2.41)	0.28	10.43 ***	(3.24)	0.71
2	3.26	(1.39)	0.02	6.29 **	(2.35)	0.34
3	2.99	(1.48)	0.04	5.32 **	(2.37)	0.32
4	2.56	(1.42)	0.02	5.34 ***	(2.65)	0.41
5	2.81 *	(1.71)	0.07	5.39 ***	(2.80)	0.49
6	2.73 *	(1.67)	0.07	5.45 ***	(2.93)	0.54
7	2.81 *	(1.79)	0.10	5.46 ***	(3.14)	0.61
8	2.24	(1.40)	0.04	5.44 ***	(3.13)	0.65
9	3.70 **	(2.18)	0.22	6.37 ***	(3.47)	0.83
Winners	3.92 *	(1.82)	0.17	6.43 ***	(2.83)	0.61

D Topic Weights Constructed by Raw Counts of Seed Words from the *NYT*

In this appendix, we conduct a robustness check for the main empirical results in the paper. Specifically, we investigate whether the sLDA model adds 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, Lang, and Stice-Lawrence (2017), Choudhury et al. (2019), Brown, Crowley, and Elliott (2020), and Bybee et al. (2023)). 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 scaling them by the total number of ngrams in the article.

Table D.1 reports the summary statistics for these topic weights. *War* is still the most frequently mentioned and most volatile topic with a monthly mean of 0.12% and standard deviation of 0.09%. It implies, on average, 0.12% of monthly *NYT* words are related to five *War* seed words (war, tension, conflict, terrorism, and terrorist).[4] This might seem low, but it shows the limitation of the raw count approach. It relies on a list of comprehensive words, and their sources can be subjective. In contrast, sLDA lets the machine captures the words co-occurring with the seed words; thus, it has less subjectivity than the words count approach. *War* has the first-order autocorrelation of 96%, 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.[5] Once again, the PLS index

[4]On average, every month, we have about 2 million ngrams in the *NYT* (the product of 4000 articles and 500 ngrams per article). For War, 0.12% of this number means 2400 mentions of the five War seed words.

[5]Comparing to the PLS weight from sLDA, the PLS weight of the seed word count is much lower due to its low topics weight. Recall that the PLS weight is the slope from regressing the topic weight on market returns; thus, the different scales of the dependent variable result in the different scale of the slope.

heavily loads on *War* and strongly correlates with this topic with a correlation coefficient of 99%.

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 the sample. Table D.2 shows that, in general, both *War* and the PLS index can powerfully and positively predict market excess returns one month ahead, consistent with the sLDA results. The other manually constructed topics, similar to the sLDA topics, do not display any consistent predictability pattern.

In Table D.3, we find that the manually constructed PLS index is not significant after controlling for specific economic predictors (book-market, long-term yield, and term spread), and, in Table D.4, the manually counted topic index loses its significance when controlling for other uncertainty variables.

The in-sample predictability results can be biased if the predictors are highly consistent, which is the case for the manually counted *War* and PLS index. Hence, in Table D.5, we report the out-of-sample R^2 computed with the frequency-based topics. Unsurprisingly, over the whole evaluation period of 1881–2019, the raw *War* index produces a much lower R^2_{OS} than the sLDA one: 0.08% versus 0.17%. The sLDA one continues to outperform in each subperiod. Similarly, the manually constructed topics 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 out-of-sample predictability. Moreover, the frequency-based topic 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 D.1
Summary Statistics
Topic Weights Constructed by Raw Counts of Seed Words

This table presents the summary statistic of the time series of 14 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) topic index, and Panel D report the correlations among topics. All numbers (except sample size) are in percentages.

	N	Mean	SD	Q1	Median	Q3	AC(1)	PLS Weights	Corr PLS
War	1784	0.12	0.09	0.07	0.09	0.12	95.78	0.10	99.29
Pandemic	1784	0.00	0.00	0.00	0.00	0.00	58.34	0.00	0.39
Panic	1784	0.05	0.02	0.03	0.04	0.05	92.92	0.02	-2.40
Confidence	1784	0.00	0.00	0.00	0.00	0.00	79.75	0.00	-1.10
Saving	1784	0.02	0.01	0.02	0.02	0.03	69.60	0.00	-0.86
Consumption	1784	0.02	0.00	0.01	0.02	0.02	73.99	0.00	14.21
Money	1784	0.12	0.05	0.09	0.12	0.14	87.86	-0.01	-51.43
Tech	1784	0.06	0.04	0.02	0.04	0.09	97.68	0.02	4.96
Real estate boom	1784	0.01	0.00	0.01	0.01	0.02	79.31	-0.00	-16.47
Real estate crash	1784	0.01	0.00	0.00	0.01	0.01	69.89	0.00	-3.77
Stock bubble	1784	0.02	0.01	0.02	0.02	0.03	68.50	-0.00	-20.57
Stock crash	1784	0.04	0.01	0.03	0.04	0.04	71.40	-0.00	-22.43
Boycott	1784	0.10	0.03	0.08	0.10	0.12	72.28	0.01	12.96
Wage	1784	0.03	0.02	0.02	0.02	0.03	84.77	0.01	26.48
PLS	1784	76.83	86.68	29.03	57.68	84.71	95.97		

Table D.2
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 topics or the PLS index constructed by raw counts of seed words, and β is the coefficient of interest that measures the strength of predictability. Returns are 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. $^*p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

	1871-2019	1871-1949	1950-2019	2000-2019
War (%)	3.18 ***	3.64 **	4.80 ***	8.07 ***
t-stat	(2.60)	(2.17)	(3.04)	(2.86)
R^2 (%)	0.26	0.23	0.82	2.15
Pandemic (%)	0.05	-0.61	0.73	6.96 ***
t-stat	(0.05)	(-0.39)	(0.54)	(2.96)
R^2 (%)	-0.06	-0.10	-0.10	1.49
Panic (%)	2.48	2.29	2.20	1.53
t-stat	(1.38)	(0.48)	(0.88)	(0.29)
R^2 (%)	0.13	0.03	0.08	-0.33
Confidence (%)	1.23	0.57	0.88	-0.45
t-stat	(1.00)	(0.22)	(0.51)	(-0.14)
R^2 (%)	-0.01	-0.10	-0.09	-0.42
Saving (%)	1.67	3.23	-0.17	5.06
t-stat	(0.93)	(0.99)	(-0.09)	(1.47)
R^2 (%)	0.03	0.16	-0.12	0.59
Consumption (%)	1.82	3.98 *	1.03	2.29
t-stat	(1.13)	(1.74)	(0.59)	(0.63)
R^2 (%)	0.05	0.29	-0.08	-0.22
Money (%)	-0.58	0.18	-1.00	-0.55
t-stat	(-0.37)	(0.08)	(-0.52)	(-0.12)
R^2 (%)	-0.05	-0.10	-0.08	-0.41
Tech (%)	1.21	-0.32	0.69	0.17
t-stat	(1.03)	(-0.12)	(0.42)	(0.05)
R^2 (%)	-0.01	-0.10	-0.10	-0.42
Real estate boom (%)	-0.68	-0.47	-3.67 *	-2.98
t-stat	(-0.47)	(-0.27)	(-1.78)	(-0.60)
R^2 (%)	-0.04	-0.10	0.43	-0.07
Real estate crash (%)	2.43 **	0.50	3.09 **	2.47
t-stat	(2.17)	(0.22)	(2.05)	(0.93)
R^2 (%)	0.12	-0.10	0.27	-0.18
Stock bubble (%)	-0.98	-1.88	-1.19	-3.38
t-stat	(-0.86)	(-1.20)	(-0.77)	(-1.07)
R^2 (%)	-0.03	-0.02	-0.06	0.03
Stock crash (%)	-0.06	-1.54	2.54	-1.23
t-stat	(-0.04)	(-0.85)	(1.22)	(-0.25)
R^2 (%)	-0.06	-0.05	0.14	-0.36
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	2.73
t-stat	(1.50)	(1.64)	(-0.10)	(0.93)
R^2 (%)	0.08	0.18	-0.12	-0.13
PLS (%)	3.13 **	3.40 **	4.78 ***	7.92 ***
t-stat	(2.54)	(2.03)	(2.93)	(2.72)
R^2 (%)	0.24	0.19	0.81	2.05
Shiller PLS (%)	1.83	0.77	2.60	4.12
t-stat	(1.43)	(0.39)	(1.49)	(1.64)
R^2 (%)	0.05	-0.09	0.16	0.25

Table D.3
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

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

in Panel A, and the following predictive regression

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

in Panel B, where R^e_{t+1} is the excess market return over the next month, x_t is the topic 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, Ringgenberg, and Zhou (2016). The last row reports the results using PLS with the 16 economic variables. Returns are 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. *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

	Panel A: Univariate		Panel B: Bivariate			
Economic Predictor	$\gamma(\%)$	$R^2(\%)$	$\beta(\%)$	$\gamma(\%)$	$R^2(\%)$	Period
Dividend-price ratio (DP)	1.39	0.00	2.99 **	0.98	0.22	187101-201910
Dividend yield (DY)	2.03	0.07	2.90 **	1.61	0.27	187102-201910
Earnings-price ratio (EP)	2.46	0.13	2.70 **	1.83	0.29	187101-201910
Dividend payout ratio (DE)	-1.00	-0.03	3.07 **	-0.74	0.21	187101-201910
Stock variance (SVAR)	-0.08	-0.06	3.13 **	-0.06	0.19	187101-201910
Book-to-market Ratio (BM)	5.19	0.58	2.40	4.80	0.63	192103-201910
Net equity expansion (NTIS)	-4.11	0.31	3.11 *	-3.91	0.45	192612-201910
Treasury bill rate (TBL)	-3.68 **	0.36	2.23 *	-3.02 *	0.44	187101-201910
Long term bond yield (LTY)	-2.82	0.11	2.50	-1.90	0.16	191901-201910
Long term bond return (LTR)	3.36 *	0.18	3.50 **	3.51 *	0.38	192601-201910
Term spread (TMS)	-2.70	0.10	2.55	-1.77	0.15	191901-201910
Default yield spread (DFY)	2.87	0.12	3.29 **	2.97	0.30	191901-201910
Default return spread (DFR)	2.30	0.04	3.34 **	2.29	0.22	192601-201910
Inflation (INFL)	-3.32	0.20	3.00 **	-3.98 *	0.34	191302-201910
Output Gap (OG)	-3.39	0.20	3.40 **	-3.65	0.40	191902-201910
Short Interest (SI)	-5.70 **	0.94	5.18 **	-6.63 ***	1.65	197301-201412
Economic PLS	4.40 *	0.48	5.85 ***	6.16 **	1.37	197301-201412

Table D.4
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 topic 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 topic 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, Ludvigson, and Ng (2015), economic policy uncertainty index from Baker, Bloom, and Davis (2016), disagreement index from Huang, Li, and Wang (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 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. *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

	Panel A: Correlations	Panel B: Univariate		Panel C: Bivariate			
Economic Predictor	Corr. with PLS (%)	γ(%)	R^2(%)	β(%)	γ(%)	R^2(%)	Period
Financial uncertainty	-7.56 **	-5.75 **	1.15	3.29 *	-5.50 *	1.43	196007-201910
Macro uncertainty	-15.02 ***	-4.30	0.58	3.13	-3.83	0.81	196007-201910
Economic policy uncertainty	17.59 ***	4.03	0.38	3.84	3.36	0.69	198501-201910
Implied volatility (VIX)	3.18	0.40	-0.27	4.93 **	0.25	0.46	199001-201910
News implied volatility (NVIX)	5.70 **	0.03	-0.07	2.90 **	-0.13	0.10	188907-201603
Disagreement	5.21	-8.43 ***	2.43	4.02 *	-8.63 ***	2.85	196912-201812
News sentiment	8.37 ***	-0.52	-0.05	3.20 **	-0.78	0.21	185701-201910
Investor sentiment (BW)	7.18 *	-2.50	0.08	3.70 *	-2.77	0.44	196507-201812
Investor sentiment (PLS)	-12.05 ***	-7.32 ***	1.86	2.66	-7.00 ***	1.97	196507-201812
Manager sentiment	-42.60 ***	-9.06 ***	3.32	2.37	-8.05 **	2.99	200301-201712
Shiller's one-year confidence index	12.50 *	-4.77	0.48	7.85 **	-5.75 **	2.54	200107-201910
Shiller's crash confidence index	16.65 **	-2.07	-0.28	7.69 ***	-3.35	1.64	200107-201910
Uncertainty PLS	-17.38 **	2.38	-0.39	6.24 *	3.46	0.59	200301-201603

Table D.5
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 topics constructed by raw counts of seed words. Panels A and B report results using OLS and PLS, respectively. All 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 1881, and the whole sample is from January 1871 to October 2019.

	1881-2019	1881-1949	1950-2019	2000-2019
Panel A: OLS				
Dividend-price ratio (DP)	-0.60	-0.81	-0.25	0.05
Dividend yield (DY)	-0.48	-0.39	-0.64	0.04
Earnings-price ratio (EP)	-0.14	-0.07	-0.26	-0.35
Dividend payout ratio (DE)	-0.83	-1.12	-0.33	-1.06
Stock variance (SVAR)	-1.68	-2.18	-0.79	-0.86
Treasury bill rate (TBL)	0.07 **	-0.05	0.26 **	0.45
War	0.08 ***	-0.02 **	0.26 *	0.77 **
Pandemic	-0.11	-0.13	-0.06	-0.41
Panic	-0.48	-0.71	-0.08	-2.37
Confidence	-0.45	-0.48	-0.39	-1.52
Saving	-0.13	-0.01	-0.34	0.36 *
Consumption	-0.20	-0.02 *	-0.52	0.11
Money	-0.17	-0.26	-0.01	-0.02
Tech	-0.34	-0.42	-0.22	-0.64
Real estate boom	-0.14	-0.13	-0.16	-0.14
Real estate crash	-0.09	-0.31	0.31 **	-1.05
Stock bubble	-0.08	-0.04	-0.14	0.12
Stock crash	-0.18	-0.17	-0.21	-0.10
Boycott	-0.26	-0.42	0.02	0.30 ***
Wage	-0.32	-0.39	-0.20	0.19
Panel B: PLS				
Economic	-0.84	-1.11	-0.38	-0.55
Topics	-0.26 *	-0.53 *	0.22 *	0.63 *
Shiller Topics	-0.69	-1.14	0.10	-0.59

E Discourse Topics from the *WSJ*

In this appendix, we check whether discourse topics extracted from 660 thousand *WSJ* articles predict stock market returns over the period 2000–2019.[6] WSJ is a mainstream media in the US, and readers are stock market participants. In contrast, *NYT*'s readers are educated and focus more on various news topics. In terms of politics, *NYT* and *WSJ* have polarized views. We apply sLDA to *WSJ* as an out-of-sample check. We focus on the past 20 years because this is the period where the *NYT* topics show the most robust predictability. We apply the same estimation method described in Section 2 to obtain the 14 time series of topic weights from the *WSJ* data.

Before extracting the 14 topics 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 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. See Appendix B for the description of text processing, cleaning, and converting into ngrams.

We plot the word clouds and time series of each topic in Figures E.1 and E.2. We report the summary statistics for these topics in Table E.1.

Table E.2 reports results in predicting the excess market returns one month ahead using all *WSJ* topics. Consistent with the *NYT* results, *War* constructed from *WSJ* is a strongly positive market predictor over 2000-2019, significant at the 1% level. Specifically, a one standard deviation increase in *War* attention is associated with an 8.3% annualized increase in market returns next month. Its R^2_{OS}, constructed in an expanding window fashion with an initial 60-month training period, is 1.53% (also significant at the 1% level). Besides *War*, *Stock Bubble* also shows significant prediction results (at the 5% level), although it is a negative predictor. Its R^2_{OS} is 0.89%, significant at the 10% level.

We also aggregate the topics with the PLS technique using all 14 topics (the "PLS" row) and 12 cases from Shiller (2019) (the "Shiller PLS" row). Both indexes display in-sample solid

[6]Following the method in Section 2, we use the first 120 months from 1990 to 1999 to construct the first monthly topic weights.

predictability. However, as the sample is small and the PLS method has many parameters to estimate, it yields poor OOS results.

Table E.3 shows prediction results over the long horizons. In line with the *NYT* results, *WSJ War* can predict the stock market returns up to 36 months ahead. *Stock Bubble* has predictability for up to 12 months, with the strongest result obtained within three months. The PLS index can strongly predict the market for up to 36 months, significant at the 1% level across all horizons. This result is expected because the PLS index is constructed to optimize its in-sample predictability over this 20-year sample.

Overall, we have found consistent results between the *NYT* and *WSJ* topics over the past 20 years. Across the two national newspapers, attention paid to the *War* topic has been a strong market predictor since 2000. In addition, we also document that *Stock Bubble* is a negative predictor for the *WSJ* articles. We conjecture that Stock Bubble captures the stock market state: news talks more about Stock Bubble when the market is overvalued, foreshadowing future corrections, resulting in pessimistic predictions.

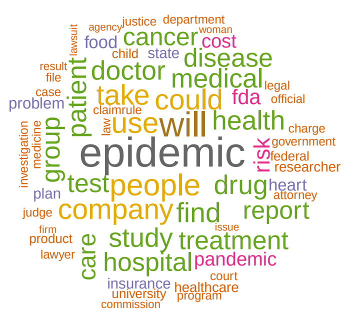
Figure E.1. Discourse Topic Contents from the *WSJ*

This figure plots the over-time frequencies of n-grams per each topic constructed according to the sLDA model. The size of each n-gram indicates its frequency. The sample period is from January 2000 to October 2019.

War

Panic

Pandemic



Monetary Standard

Real Estate Booms

Boycotts and Evil Business

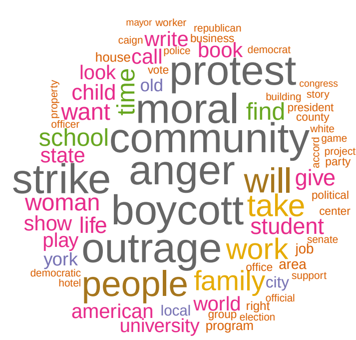


Figure E.1. Discourse Topic Contents from the *WSJ* (Cont.)

This figure plots the over-time frequencies of n-grams per each topic constructed according to the sLDA model. The size of each n-gram indicates its frequency. The sample period is from January 2000 to October 2019.

Figure E.2. Time Series of Topic 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 January 2000 to October 2019.

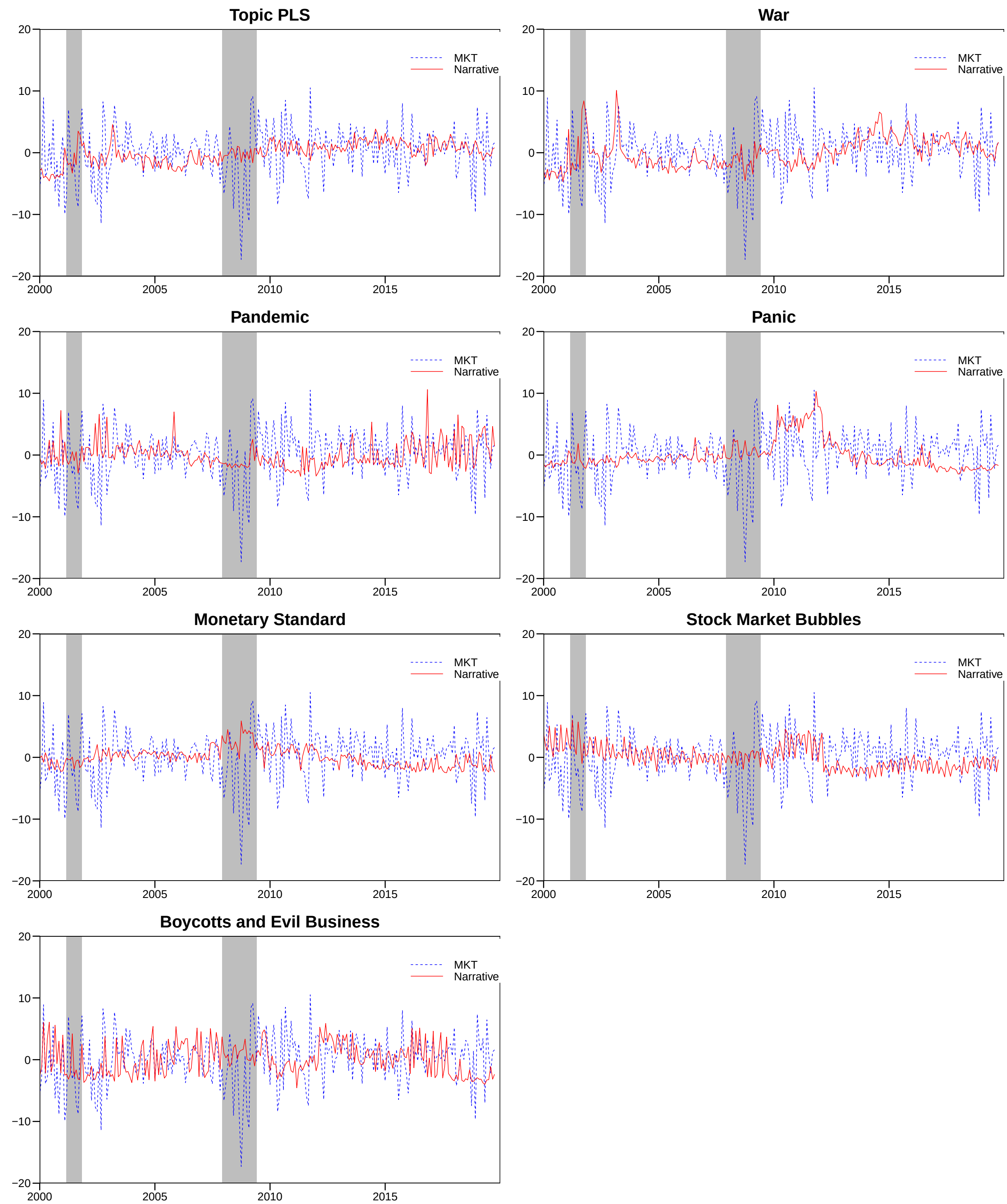


Figure E.2. Time Series of Topic Weights from the *WSJ* (Cont.)

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 January 2000 to October 2019.

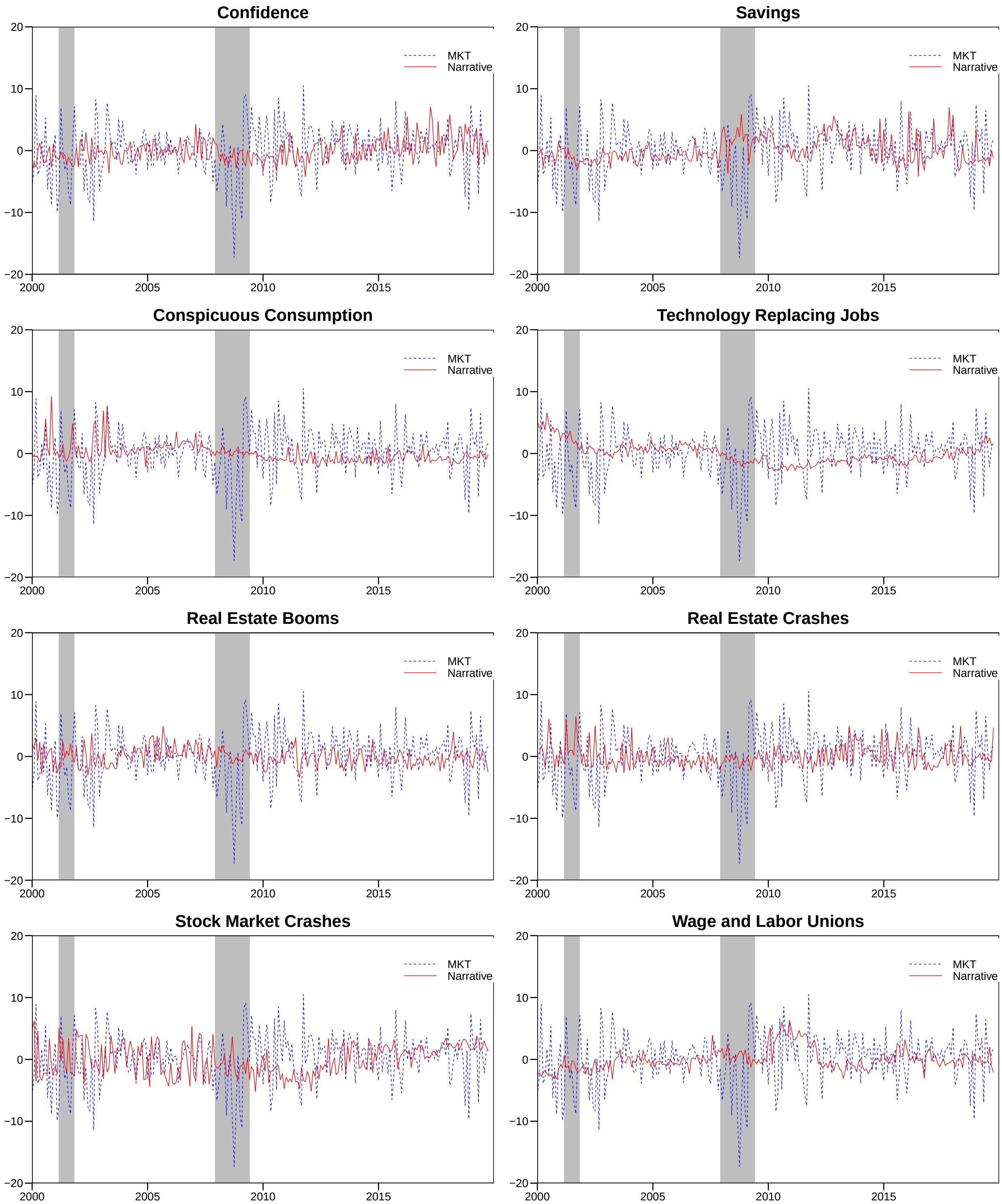


Table E.1
Summary Statistics of Topics from the *WSJ*

This table presents the summary statistics for the time series of 14 monthly topic weights constructed from the *WSJ* articles according to the sLDA model described in Section 2. All numbers (except sample size) are expressed as percentages. The sample period is from January 2000 to October 2019.

	N	Mean	SD	Q1	Median	Q3	AC(1)	PLS Weights	Corr PLS
War	238	8.87	2.42	7.13	8.61	10.21	70.54	9.48	77.05
Pandemic	238	5.77	2.16	4.35	5.31	6.74	14.11	-4.32	-23.65
Panic	238	5.95	2.40	4.47	5.30	6.34	87.48	2.54	11.90
Confidence	238	6.39	1.88	5.04	6.08	7.58	15.07	-1.19	12.53
Saving	238	7.60	2.04	6.23	7.11	8.68	39.02	2.90	24.23
Consumption	238	5.14	1.52	4.14	4.87	5.71	30.22	-0.11	-27.67
Money	238	6.59	1.52	5.46	6.48	7.34	67.08	-1.14	-17.30
Tech	238	6.41	1.66	5.21	6.19	7.32	87.43	-3.28	-63.22
Real estate boom	238	6.12	1.47	5.08	6.03	7.01	16.16	-2.35	-31.04
Real estate crash	238	5.40	1.74	4.22	4.96	6.23	2.45	0.56	4.64
Stock bubble	238	5.72	1.91	4.33	5.59	6.67	43.85	-6.47	-43.81
Stock crash	238	8.01	2.54	5.85	8.05	10.07	31.80	1.41	23.18
Boycott	238	8.65	2.65	6.32	8.27	10.48	24.26	-4.90	-30.75
Wage	238	6.91	1.75	5.90	6.63	7.44	72.83	3.32	28.17
PLS	238	12.17	17.43	2.07	13.76	24.30	66.35		

Table E.2
Predicting One-Month Market Returns with *WSJ* Topics

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 topics or the PLS indexes constructed from the *WSJ* articles, and β, the coefficient of interest, measures the strength of predictability. "Shiller PLS" uses only the topics from Shiller (2019), excluding *War* and Pandemic. 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 out-of-sample R^2 (R^2_{OS}) is computed using an expanding window with the initial estimation window of 60 months and is evaluated based on the Clark and West (2007) MSFE-adjusted statistic. The sample is from January 2000 to October 2019. *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

	β (%)	t-stat	R^2 (%)	R^2_{OS} (%)
War	8.31 ***	(2.75)	2.30	1.53 ***
Pandemic	-4.24	(-1.24)	0.29	-2.63
Panic	2.24	(0.68)	-0.23	-2.64
Confidence	-1.34	(-0.48)	-0.35	-2.42
Saving	3.02	(0.86)	-0.06	-1.98
Consumption	-0.15	(-0.05)	-0.42	-3.31
Money	-1.60	(-0.38)	-0.32	-4.42
Tech	-4.19	(-1.23)	0.27	-1.02
Real estate boom	-3.40	(-1.07)	0.03	-1.36
Real estate crash	0.68	(0.27)	-0.41	-0.66
Stock bubble	-7.17 **	(-2.14)	1.61	0.89 *
Stock crash	1.17	(0.35)	-0.37	-0.67
Boycott	-3.93	(-1.45)	0.18	-0.85
Wage	4.03	(1.15)	0.22	-1.46
PLS	12.17 ***	(4.52)	5.42	-3.58
Shiller PLS	9.86 ***	(3.10)	3.41	-7.27

Table E.3
Predicting Long-Horizon Market Returns with *WSJ* Topics

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 either *War*, Stock Bubble or the PLS index constructed from the *WSJ* articles, 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 is from January 2000 to October 2019. *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

	War (%)	t-stat	R^2 (%)	Stock Bubble (%)	t-stat	R^2 (%)	PLS (%)	t-stat	R^2 (%)	N
$h = 1$	8.31 ***	(2.75)	2.30	-7.17 **	(-2.14)	1.61	12.17 ***	(4.52)	5.42	238
$h = 3$	6.94 ***	(3.26)	4.93	-5.63 ***	(-2.77)	3.10	8.49 ***	(4.37)	7.59	238
$h = 6$	4.45 **	(2.32)	3.43	-4.71 **	(-2.22)	3.89	6.20 ***	(3.63)	7.03	238
$h = 12$	3.68	(1.49)	4.25	-4.53 **	(-2.19)	6.64	5.68 ***	(2.76)	10.70	238
$h = 24$	3.70 **	(2.01)	7.49	-3.02	(-1.35)	4.83	6.01 ***	(3.27)	20.50	230
$h = 36$	3.77 **	(2.24)	11.80	-1.85	(-0.89)	2.51	6.45 ***	(4.82)	35.47	218