

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 <i>NYT</i>	7
C.1 <i>War</i> versus Conditional Volatility and Skewness	7
C.2 <i>War</i> as a Proxy for Time-Varying Risk Aversion	8
C.3 <i>War</i> 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 <i>NYT</i>	26
E Discourse Topics from the <i>WSJ</i>	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

¹To 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.

²Similarly, 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:

$$z_{dv} \sim \text{Multinomial}(\tau_d) \text{ (a distribution which does not depend on } v\text{), and then}$$
 - 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.

³Gibbs 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:

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

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:

$$R_{t+1}^e = \alpha + \beta War_t + \gamma z_t + \epsilon_{t+1}, \quad (\text{C.1})$$

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:

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

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:

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

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, $\sigma_{M,t}^2$, and its conditional covariance with the investment opportunity set, $\sigma_{MF,t}$:

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

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):

$$\mathbb{E}_t[R_{M,t+1} - R_{f,t+1}] = \lambda_0 + \lambda_1 \times \sigma_{M,t}^2, \quad (\text{C.5})$$

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 by Bollerslev (1986) and Engle and Bollerslev (1986). Specifically, we consider first the following mean equation for the return-volatility trade-off:

$$R_{M,t+1} - R_{f,t+1} = \lambda_0 + \lambda_1 \times \sigma_{M,t}^2 + \epsilon_{t+1}, \quad (\text{C.6})$$

where ϵ_{t+1} has a mean of zero with conditional variance $\sigma_{M,t}^2$. 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 :

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

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):

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

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:

$$R_{t+1}^e = \alpha + \beta \times War_t + \gamma^j \times D_t^j + \epsilon_{t+1}, \quad (\text{C.9})$$

where D_t^j 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:

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

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:

$$R_{i,t+1}^e = \alpha_i + \beta_i x_t + \epsilon_{i,t+1}, \quad i = 1, \dots, 40 \quad (\text{C.11})$$

where $R_{i,t+1}^e$ 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.

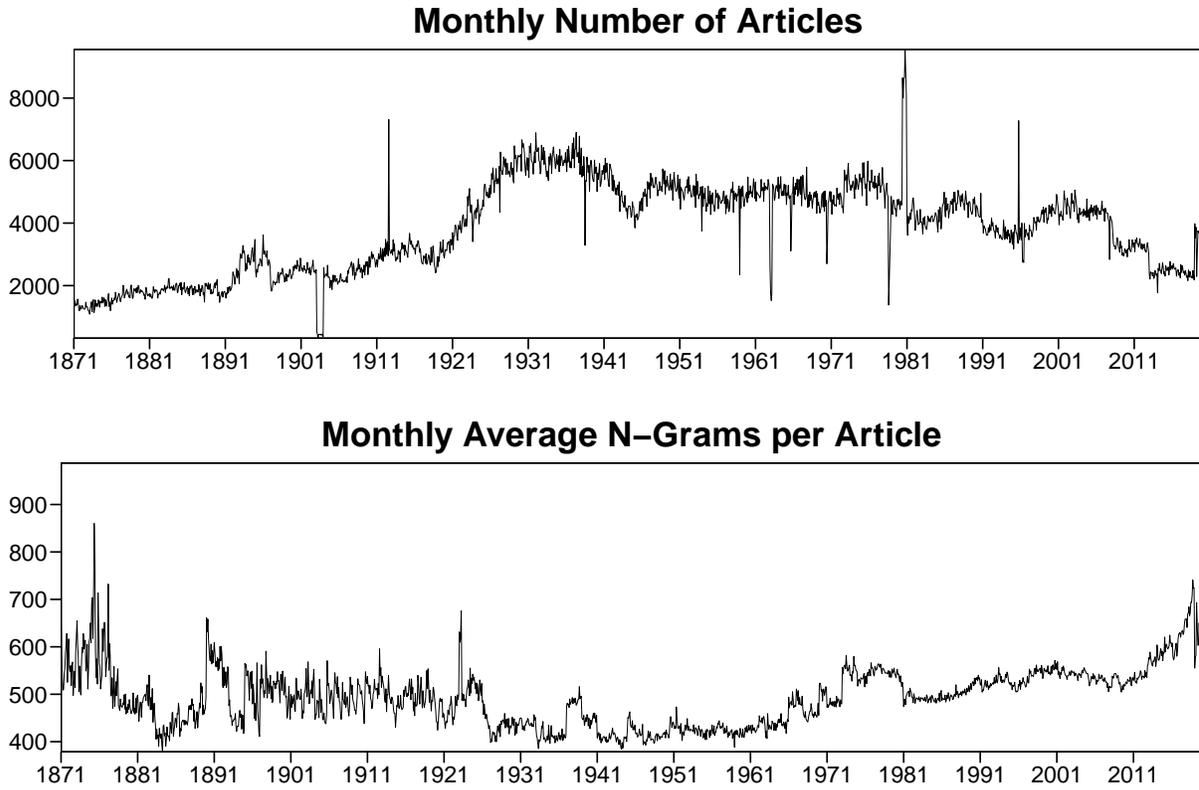


Figure C.3. Time Series of Topics Weights from the *NYT*

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.

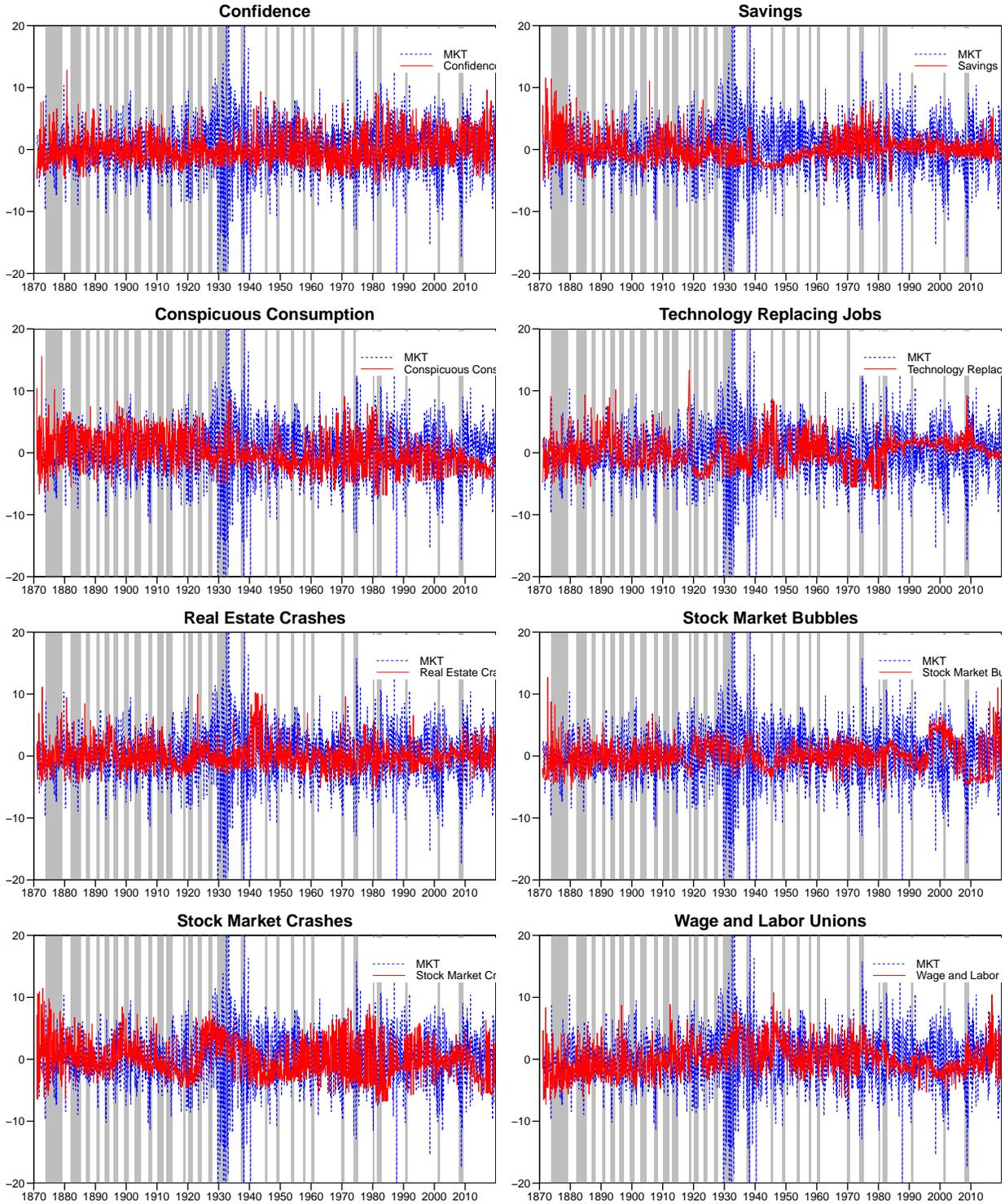


Figure C.4. Historical Events

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.

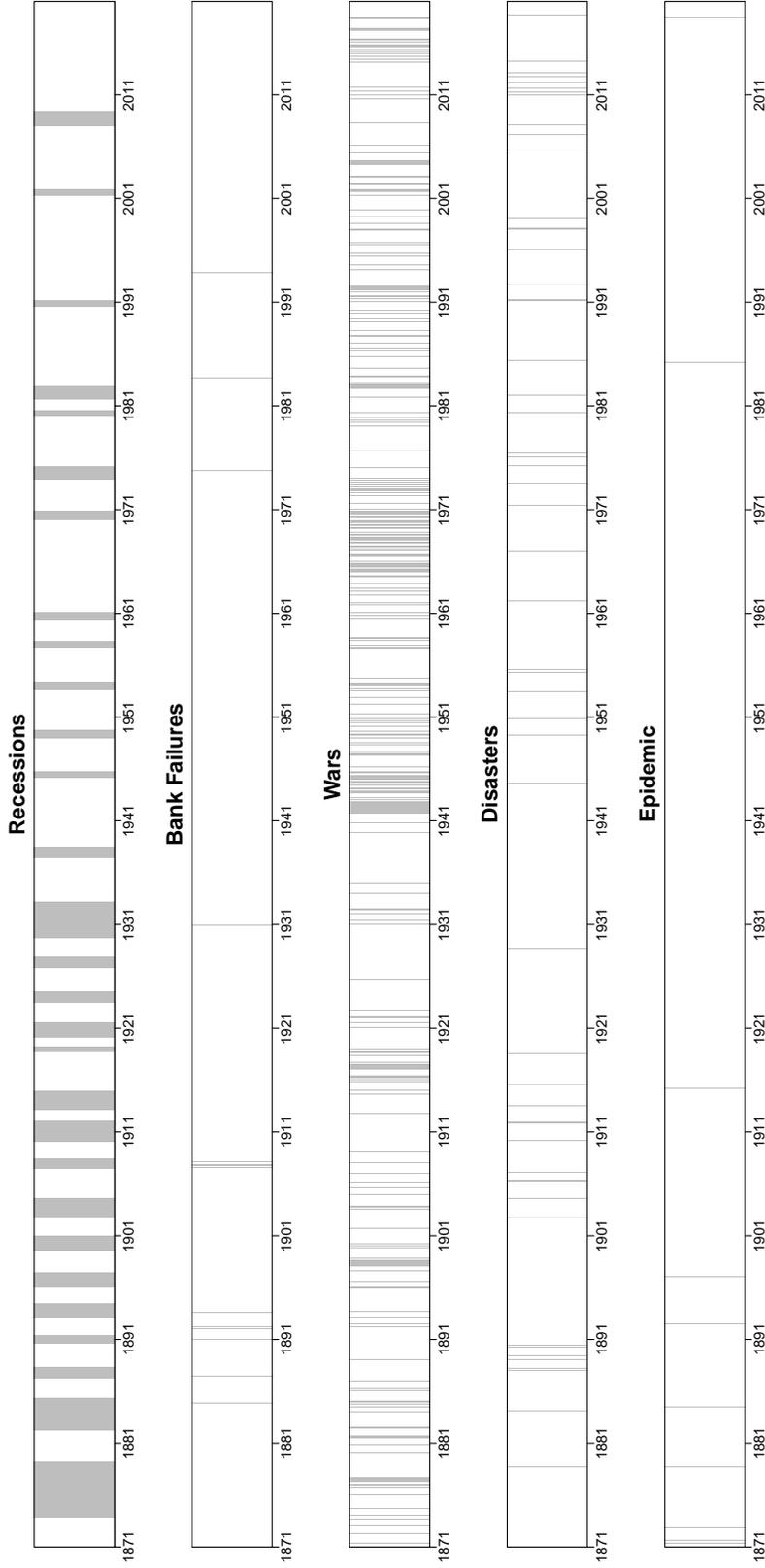


Table C.1
Data Screening

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															
Pandemic	-6.5														
Panic	-17.1	-6.5													
Confidence	-7.9	-10.7	-6.2												
Saving	-22.2	-2.2	-14.9	-3.3											
Consumption	-26.2	-8.8	5.3	-10.0	-3.3										
Money	-10.0	-2.3	5.3	-10.0	-5.8	10.4									
Tech	-1.5	-4.5	-8.2	6.6	-4.9	10.4	-14.6								
Real estate boom	-1.7	-7.3	-6.1	-8.5	-6.8	-14.4	-14.6	-4.1							
Real estate crash	-10.1	-1.2	-3.0	-1.6	-7.3	-4.3	-12.4	-2.8	-1.8						
Stock bubble	-10.1	-3.7	-15.3	-5.9	-0.8	-7.0	-10.8	-8.6	-0.9	-7.7					
Stock crash	-19.9	-13.9	-14.1	-20.0	3.1	-1.2	1.8	-7.0	-17.4	-6.0	-0.0				
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			
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		
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:

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

where R_{t+1}^e 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 ** (2.13)	3.71 ** (2.27)	3.48 ** (2.16)	3.61 ** (2.27)
MKT	5.28 (1.39)			6.22 (1.46)
VOL		0.68 (0.14)		2.31 (0.44)
SK			-2.33 (-0.89)	-3.27 (-1.37)
R^2	0.78	0.13	0.25	0.93
Panel B: 1950-2019				
War	3.97 ** (2.54)	3.71 ** (2.33)	4.00 ** (2.53)	3.63 ** (2.30)
MKT	1.44 (0.66)			1.10 (0.53)
VOL		-1.92 (-0.67)		-1.62 (-0.57)
SK			-1.29 (-0.74)	-1.45 (-0.81)
R^2	0.52	0.58	0.50	0.46
Panel C: 2000-2019				
War	9.49 *** (3.31)	8.81 *** (2.92)	9.75 *** (3.39)	8.71 *** (2.86)
MKT	2.27 (0.55)			0.36 (0.09)
VOL		-5.10 (-0.95)		-4.94 (-0.86)
SK			-1.01 (-0.38)	-0.99 (-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):

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

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

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

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):

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

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:

$$R_{t+1}^e = \alpha + \beta \times War_t + \gamma^j \times D_t^j + \epsilon_{t+1}$$

where R_{t+1}^e is the excess market return over the next month, D_t^j 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						
<i>War</i>	3.02 ** (2.52)	3.82 *** (3.37)	3.58 *** (3.06)	3.75 *** (3.32)	3.84 *** (3.40)	3.55 *** (3.14)
Event	-8.33 ** (-2.08)	3.36 (0.30)	2.90 (0.72)	-7.06 (-1.10)	18.32 * (1.66)	-8.81 *** (-3.07)
R^2 (%)	0.74	0.33	0.37	0.38	0.39	0.92
Panel B: 1871-1949						
<i>War</i>	2.54 (1.37)	3.62 ** (2.09)	2.68 (1.45)	3.42 ** (1.98)	3.54 ** (2.05)	3.35 * (1.93)
Event	-11.24 ** (-2.29)	12.69 (1.07)	9.36 (1.32)	-8.11 (-1.15)	20.43 (1.57)	-11.65 *** (-2.85)
R^2 (%)	0.85	0.14	0.38	0.14	0.19	0.95
Panel C: 1950-2019						
<i>War</i>	4.12 ** (2.58)	4.02 ** (2.53)	4.16 *** (2.62)	4.01 ** (2.53)	4.07 ** (2.57)	4.26 *** (2.66)
Event	-4.55 (-0.64)	-32.15 * (-1.80)	-2.44 (-0.56)	-6.22 (-0.64)	9.74 (0.51)	-6.19 (-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_{exp}^2	R_{rec}^2	R_{high}^2	R_{low}^2
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:

$$R_{i,t+1}^e = \alpha_i + \beta_i x_t + \epsilon_{i,t+1}, \quad i = 1, \dots, 40$$

where $R_{i,t+1}^e$ 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. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

	<i>War</i> (%)	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. (2021)). This paper blends the two branches by employing a semisupervised model in which we inject seed words into the topic model to extract desired contents. Hence, a natural question is whether the sLDA model reveals any additional information beyond a simple count of those seed words in the news. To answer this question, we construct topic weights by simply counting the occurrences of seed words and 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

⁴On 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.

⁵Comparing 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_{OS}^2 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
<i>War</i>	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:

$$R_{t+1}^e = \alpha + \beta x_t + \epsilon_{t+1 \rightarrow t+1},$$

where R_{t+1}^e 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
<i>War</i> (%)	3.18 ***	3.64 **	4.80 ***	8.07 ***
<i>t</i> -stat	(2.60)	(2.17)	(3.04)	(2.86)
R^2 (%)	0.26	0.23	0.82	2.15
<i>Pandemic</i> (%)	0.05	-0.61	0.73	6.96 ***
<i>t</i> -stat	(0.05)	(-0.39)	(0.54)	(2.96)
R^2 (%)	-0.06	-0.10	-0.10	1.49
<i>Panic</i> (%)	2.48	2.29	2.20	1.53
<i>t</i> -stat	(1.38)	(0.48)	(0.88)	(0.29)
R^2 (%)	0.13	0.03	0.08	-0.33
<i>Confidence</i> (%)	1.23	0.57	0.88	-0.45
<i>t</i> -stat	(1.00)	(0.22)	(0.51)	(-0.14)
R^2 (%)	-0.01	-0.10	-0.09	-0.42
<i>Saving</i> (%)	1.67	3.23	-0.17	5.06
<i>t</i> -stat	(0.93)	(0.99)	(-0.09)	(1.47)
R^2 (%)	0.03	0.16	-0.12	0.59
<i>Consumption</i> (%)	1.82	3.98 *	1.03	2.29
<i>t</i> -stat	(1.13)	(1.74)	(0.59)	(0.63)
R^2 (%)	0.05	0.29	-0.08	-0.22
<i>Money</i> (%)	-0.58	0.18	-1.00	-0.55
<i>t</i> -stat	(-0.37)	(0.08)	(-0.52)	(-0.12)
R^2 (%)	-0.05	-0.10	-0.08	-0.41
<i>Tech</i> (%)	1.21	-0.32	0.69	0.17
<i>t</i> -stat	(1.03)	(-0.12)	(0.42)	(0.05)
R^2 (%)	-0.01	-0.10	-0.10	-0.42
<i>Real estate boom</i> (%)	-0.68	-0.47	-3.67 *	-2.98
<i>t</i> -stat	(-0.47)	(-0.27)	(-1.78)	(-0.60)
R^2 (%)	-0.04	-0.10	0.43	-0.07
<i>Real estate crash</i> (%)	2.43 **	0.50	3.09 **	2.47
<i>t</i> -stat	(2.17)	(0.22)	(2.05)	(0.93)
R^2 (%)	0.12	-0.10	0.27	-0.18
<i>Stock bubble</i> (%)	-0.98	-1.88	-1.19	-3.38
<i>t</i> -stat	(-0.86)	(-1.20)	(-0.77)	(-1.07)
R^2 (%)	-0.03	-0.02	-0.06	0.03
<i>Stock crash</i> (%)	-0.06	-1.54	2.54	-1.23
<i>t</i> -stat	(-0.04)	(-0.85)	(1.22)	(-0.25)
R^2 (%)	-0.06	-0.05	0.14	-0.36
<i>Boycott</i> (%)	1.15	1.01	0.22	7.29 ***
<i>t</i> -stat	(0.86)	(0.56)	(0.13)	(2.96)
R^2 (%)	-0.02	-0.08	-0.12	1.67
<i>Wage</i> (%)	2.13	3.34	-0.21	2.73
<i>t</i> -stat	(1.50)	(1.64)	(-0.10)	(0.93)
R^2 (%)	0.08	0.18	-0.12	-0.13
<i>PLS</i> (%)	3.13 **	3.40 **	4.78 ***	7.92 ***
<i>t</i> -stat	(2.54)	(2.03)	(2.93)	(2.72)
R^2 (%)	0.24	0.19	0.81	2.05
<i>Shiller PLS</i> (%)	1.83	0.77	2.60	4.12
<i>t</i> -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

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

in Panel A, and the following predictive regression

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

in Panel B, where R_{t+1}^e is the excess market return over the next month, x_t is the 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. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Economic Predictor	Panel A: Univariate		Panel B: Bivariate			Period
	$\gamma(\%)$	$R^2(\%)$	$\beta(\%)$	$\gamma(\%)$	$R^2(\%)$	
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

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

in Panel B, and the following predictive regression

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

in Panel C, where R_{t+1}^e 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. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Economic Predictor	Panel A: Correlations	Panel B: Univariate		Panel C: Bivariate			Period
	Corr. with PLS (%)	γ (%)	R^2 (%)	β (%)	γ (%)	R^2 (%)	
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_{OS}^2 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_{OS}^2 , 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_{OS}^2 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

⁶Following 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.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.

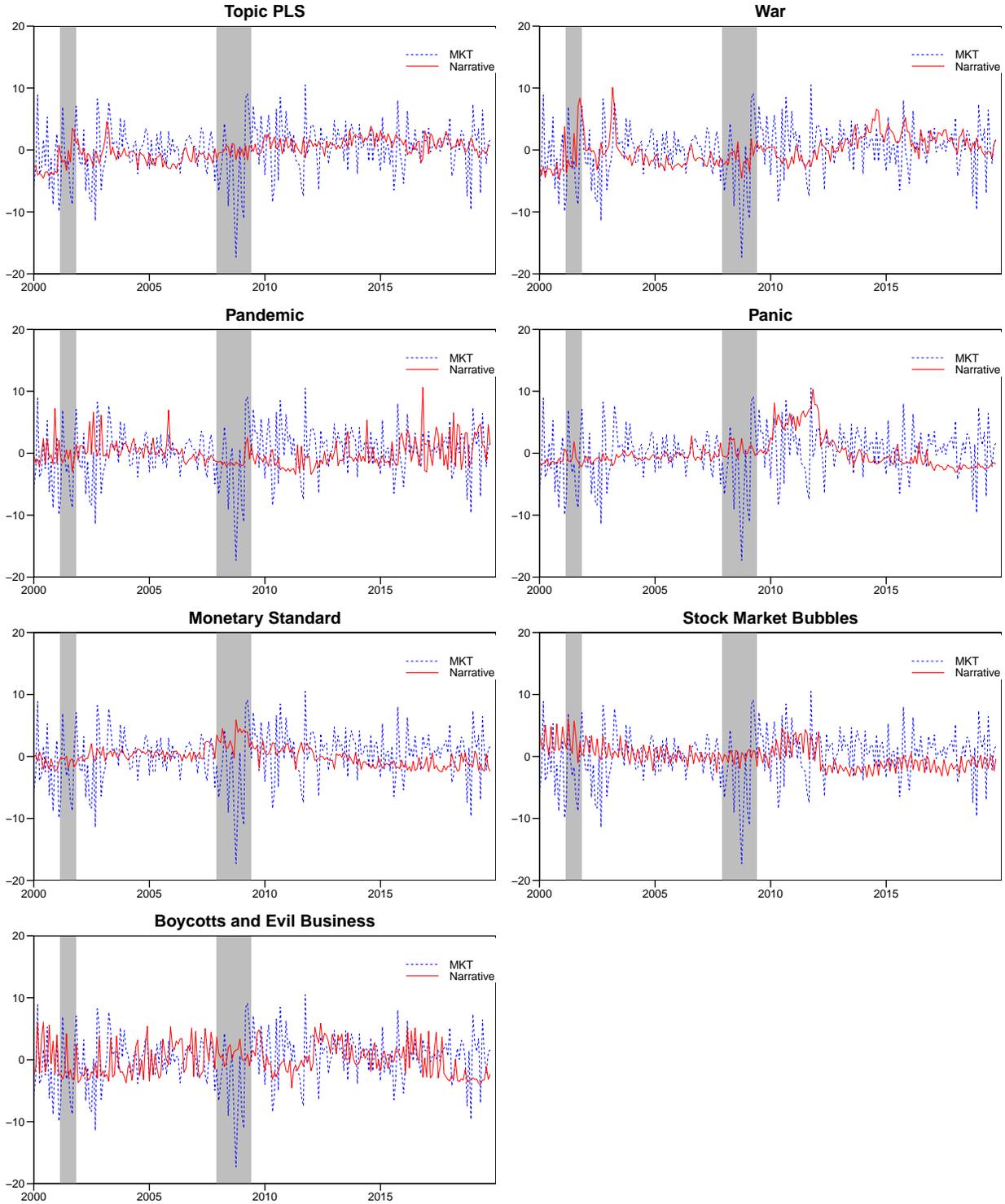


Figure E.2. Time Series of Topic Weights from the *WSJ* (Cont.)

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.

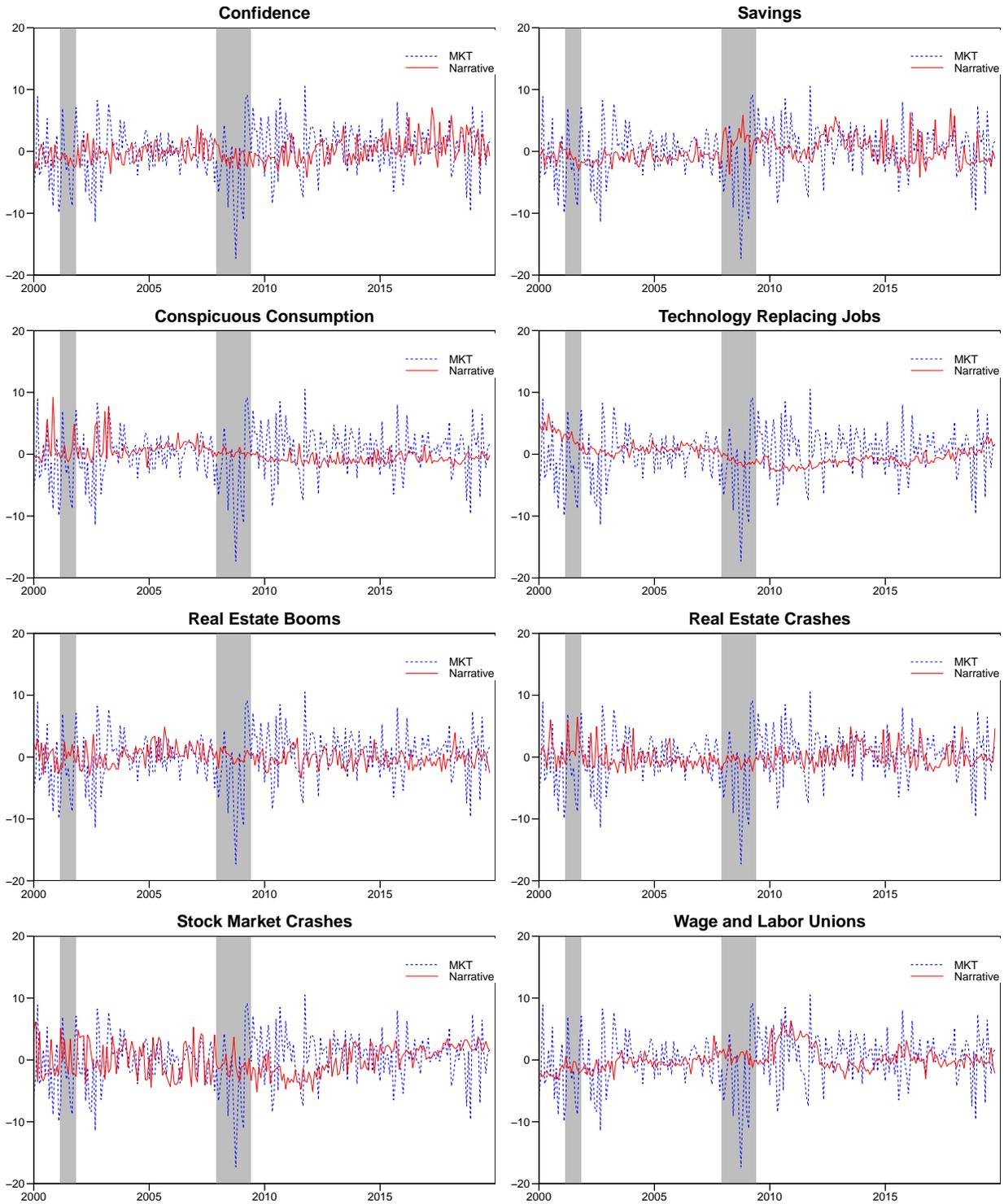


Table E.1
Summary Statistics of Topics from the *WSJ*

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
<i>War</i>	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:

$$R_{t+1}^e = \alpha + \beta x_t + \epsilon_{t+1 \rightarrow t+1},$$

where R_{t+1}^e 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_{OS}^2) 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_{OS}^2 (%)
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:

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

where $R_{t+1 \rightarrow t+h}^e$ 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