The Journal of FINANCE
The Journal of THE AMERICAN FINANCE ASSOCIATION

THE JOURNAL OF FINANCE • VOL. LXXIX, NO. 5 • OCTOBER 2024

Business News and Business Cycles

LELAND BYBEE, BRYAN KELLY, ASAF MANELA, and DACHENG XIU*

ABSTRACT

We propose an approach to measuring the state of the economy via textual analysis of business news. From the full text of 800,000 *Wall Street Journal* articles for 1984 to 2017, we estimate a topic model that summarizes business news into interpretable topical themes and quantifies the proportion of news attention allocated to each theme over time. News attention closely tracks a wide range of economic activities and can forecast aggregate stock market returns. A text-augmented vector autoregression demonstrates the large incremental role of news text in forecasting macroeconomic dynamics. We retrieve the narratives that underlie these improvements in market and business cycle forecasts.

A CENTRAL FUNCTION OF THE economics profession is measuring the state of the economy and developing models that link these measurements to distributions of future outcomes. This arms consumers, investors, and policymakers with the information and structural context necessary to allocate resources efficiently. But the economy is a complex system whose current state defies simple measurement: vast resources, both public and private, are devoted to measuring the many facets of economic activity.[1] Researchers routinely

*Leland Bybee is at Yale University. Bryan Kelly is at Yale University, AQR, and NBER. Asaf Manela is at Washington University in St. Louis and Reichman University. Dacheng Xiu is at University of Chicago. Previously titled "The Structure of Economic News." We benefited from discussions and comments from George-Marios Angeletos; Scott Baker; Nick Bloom; Steve Davis; Rob Engle; Xavier Gabaix; Will Goetzmann; Toby Moskowitz; Lasse Pedersen; Jesse Shapiro; Bob Shiller; Alp Simsek; and participants at many seminars and conferences. AQR Capital Management is a global investment management firm, which may or may not apply similar investment techniques or methods of analysis as described herein. The views expressed here are those of the authors and not necessarily those of AQR. We have read *The Journal of Finance* disclosure policy and have nothing further to disclose.

Correspondence: Asaf Manela, Olin Business School, Washington University in St. Louis, One Brookings Drive, Saint Louis, MO 63130; e-mail: amanela@wustl.edu.

[1] For example, the Bureau of Economic Analysis maintains detailed national accounts for income and spending at the aggregate and industry level, the Bureau of Labor Statistics specializes in measuring employment, wages, and general price levels, and the Federal Reserve disseminates large data sets of macroeconomic indicators (McCracken and Ng (2016)). In addition, equity and debt markets provide a barometer for economic conditions facing the corporate sector, and

DOI: 10.1111/jofi.13377

analyze this overlapping ecosystem of numerical business cycle indicators. Yet ascertaining the nature and evolution of the state of the economy from these data is a notoriously difficult task.

In this paper, we offer an approach to measuring the state of the economy via textual analysis of business news. The media sector, as a central information intermediary, continually transforms perceptions of economic events into a verbal description that we call "news." This transformation involves describing events, interpreting their meaning, forecasting their impact, and inferring their causes. The information disseminated by news media is an equilibrium outcome determined by the confluence of consumer preferences and news production technologies (Mullainathan and Shleifer (2005)). As such, news text is a mirror of the prevailing economic issues that are important to both news consumers and producers. News is thus a one-stop shop for simultaneously understanding many facets of the state of the economy and how they interact with each other.

While the media reflects information that consumers rely on to make allocation decisions in an evolving environment, little work has directly studied the structure of news. We focus our analysis on the full text of the *Wall Street Journal* (*WSJ*), over the 1984 to 2017 period, which consists of approximately 800,000 articles. We summarize this dense verbal description of the state of the economy via a topic model called latent Dirichlet allocation (LDA). Topic models are a popular dimension-reduction technique from the fields of machine learning and natural language processing. They have two essential elements. Just as principal component analysis (PCA) condenses large data matrices into a comparatively low number of common factors, a topic model's first element reduces an inherently ultrahigh-dimension representation of a text corpus into a relatively low-dimensional set of common "topics." The formation of topics is unsupervised—they are estimated as clusters of terms (words or phrases) that tend to co-occur in articles. Those clusters are optimized so that relatively few clusters (many fewer than the number of distinct terms in the data set) preserve as much of the meaning in the original corpus as possible by best explaining the variation in term usage across articles.

A topic model's second main element estimates the proportion of text dedicated to each topic in an article. These proportions are a valuable map from common topics to individual, article-level narratives that invoke those topics. More importantly, they quantify the amount of news *attention* allocated to each topic. This makes it possible, for example, to analyze the interaction between news and economic activity. In sum, our topic model accomplishes the two important tasks of (i) summarizing dominant themes throughout the history of the *WSJ* and (ii) tracking how media attention to news topics evolves. We study these two elements as a new quantitative description of the state of the economy.

academic researchers maintain and distribute indices of economic volatility (Engle (2019)) and recession probabilities (Chauvet and Piger (2008)).

Our main empirical findings are the following. First, we characterize the topical structure in business news. *WSJ* news decomposes into easily interpretable topics with intuitive time-series patterns. A model with 180 topics is a statistically optimal specification according to a Bayes factor criterion. Models with fewer topics tend to mix news themes into overly broad clusters, while allowing for more topics adds parameters without reliably improving model fit.

Almost all topics exhibit strong time-series persistence. A long-standing question in financial research is why asset returns exhibit such strong volatility clustering. While price volatility could help drive news coverage, it is unclear why volatility would persist. A leading hypothesis for this empirical fact is that volatility is driven by news arrivals, and that news itself arrives in clusters (e.g., Engle, Ito, and Lin (1990)). Newspaper coverage is naturally persistent because journalists pay a fixed cost to learn a new topic and then write cheaper follow-up stories (Boydstun (2013)). The persistence in news topic attention supports this hypothesis. For example, *WSJ* coverage of news associated with the "oil drilling" topic {key terms: *exxon mobil, cubic feet, drill rig, offshore oil*} and "oil market" topic {key terms: *opec, nonopec, oil minister, oil demand*} has similar time-series dynamics as the volatility of crude oil prices.

We find that news is a combination of recurrent, seasonal, and emergent topics. Recurrent topics are those that receive media attention consistently throughout the sample. For example, the "Federal Reserve" topic {key terms: *greenspan, yellen, federalfunds rate, fomc*} and the "health insurance" topic {key terms: *hmo, health plan, health coverage, blue cross*} are active throughout the sample. Seasonal topics include "presidential elections" {key terms: *obama, romney, dukakis, campaign finance*} and "earnings forecasts" {key terms: *analyst poll, earn forecast, earningspershare, earn expectation*}[2] draw attention with cyclical regularity. Emergent topics do not appear in much of the sample, are triggered by particular events, and often remain elevated thereafter. Examples are the "terrorism" topic {key terms: *taliban, queda, suicide bomber, osama*} and the "natural disasters" topic {key terms: *katrina, quake, tsunami, hurricane*}. These examples also illustrate that many topics describe the subjects of news (elections, the Fed, and earnings), and do not assign an obvious value of assessment of good news versus bad news. For other topics, the subject itself carries a value assignment (terrorism, recession, and natural disasters). A small number of sentiment-related topics, such as the "concerns" topic {key terms: *raise concern, major concern, express concern, increase concern*}, provide directional color to the subject topics that co-occur in a given article.

We report a variety of validation checks that show that economic topics identified from the *WSJ* text coincide with conceptually related measures of specific economic activities. For many data series, ranging from output and

[2] The phrasing of key terms reflects text processing choices described in the Internet Appendix, Section IA.A. The Internet Appendix may be found in the online version of this article. For example, "earn forecast" and "earn expectation" reflect stemming and lemmatization, which reduce "earnings" to "earn" and "expectations" to "expectation." Likewise, "earningspershare" is an example of the treatment of punctuation (dropping hyphens in this case).

employment to financing activity, asset prices, and uncertainty, we find that a small subset of thematically related news topics provide a close match to the paths of numerical macro series. Remarkably, we find that news attention explains 25% of the variation in aggregate stock market fluctuations.

More detailed examples of economic activity further demonstrate the explanatory power of news attention. We find that the volume of leveraged buyout (LBO) transactions is most associated with the "takeovers" topic {key terms: *poison pill, hostile takeover, share tender, higher offer*} and "control stakes" topic {key terms: *control stake, majority stake, minority shareholder, acquire stake*}, which help explain 58% of the variation in LBO activity. Similarly, the time-series behavior of initial public offering (IPO) volume is closely tracked by the attention to the "IPO" topic {key terms: *ipo market, ipo price, roadshow, lockup*} and "venture capital" topic {key terms: *joint venture, venture capitalist, venture fund*}. The key term lists of statistically related topics offer an immediate narrative for each numerical time series that we study. These are not causal narratives. In some cases they may reflect proximate causes, and in other cases they may represent anticipated impacts.

Our second contribution is to show that news text contains distinct, incremental information compared to standard numerical indicators. We consider two important applications of our topic model: macroeconomic vector autoregression (VAR) and stock market timing. First, we study news attention within an otherwise standard macroeconomic VAR. We find that news attention to the "recession" topic has economically large and highly significant predictive power for future output and employment, after controlling for a variety of common VAR components such as stock prices, interest rates, and measures of economic uncertainty. That is, news coverage of recession risk captures useful information about future economic outcomes above and beyond commonly used indicators. This relationship is robust and continues to hold in a much longer sample stretching back to 1890 for which we have only *WSJ* front-page text.

Going beyond a single news topic in the VAR poses an overfitting challenge, which we tackle. We seek a VAR that includes any news topics that genuinely influence macroeconomic dynamics and avoids spurious inclusion of irrelevant topics. This is a model selection problem, which we solve with cross-validated group-lasso regression. Our "text-augmented VAR" uses the group-lasso (Yuan and Lin (2006)) to select or remove an explanatory variable in its entirety (rather than selecting, say, only the third lag of a given variable). We find that the optimal cross-validated model selects a single news attention topic, "recession," for inclusion in the VAR.

Modeling economic activity in terms of news-based topics offers economists a new set of tools for understanding the drivers of economic fluctuations. In a standard macroeconomic VAR, interpretation often boils down to a choice of rotation for the error covariance matrix and the resulting impulse responses. With a text-augmented VAR, fluctuations can be directly mapped to textual narratives of economic conditions. We describe an approach for *narrative retrieval* that combines estimated VAR coefficients with topic model estimates and does not require identifying restrictions or economic constraints. It locates the most influential individual articles—that is, specific narratives—for interpreting the behavior of model-based expectations. For example, we find that large shifts in output growth expectations are associated with news article headlines such as “Consumer Confidence Slides on Fears of Layoffs” and “Stocks Fell Further Amid Concerns Prices Don’t Fully Reflect Worsening Global Growth.” Using the topic model as a narrative retrieval device, the researcher can map changes in model forecasts to nuanced textual interpretations of economic events (i.e., news articles). In essence, the estimated model flags articles that the researcher should read thoroughly, without requiring a close manual read of every article in the *WSJ*.

In a second application, we investigate whether news text, as summarized by our topic model, helps detect time variation in the equity risk premium (i.e., the expected excess return of the stock market). Following standard approaches in the literature, we estimate time-varying expected returns with return forecasting regressions. Forecasting with topic attention requires care to avoid look-ahead bias that might arise from LDA (which is estimated from the full sample of text). To avoid such bias, we use Online LDA (oLDA) (Hoffman, Bach, and Blei (2010)), which processes documents in sequential batches. This means that oLDA topic model estimates at a point in time are based only on the text observed up to that point, thereby avoiding any look-ahead bias that could pollute subsequent return forecasting regressions.

We conduct out-of-sample return forecasting analysis using LDA topic attention series as predictors. We evaluate the statistical and economic significance of these predictions through the lens of a market-timing strategy (Campbell and Thompson (2008), Kelly, Malamud, and Zhou (2024)). The out-of-sample market-timing strategy significantly outperforms several natural benchmarks including a buy-and-hold market strategy, a strategy inspired by *The Economist*’s R-word index, a strategy based on changes in economic policy uncertainty (EPU), and two strategies based on the Welch and Goyal (2008, henceforth GW) predictors. Even though the LDA topics are chosen in an unsupervised manner to reduce the dimensionality of newspaper text, this reduction is so effective at summarizing the information content of news that it generates reliable market return forecasts.

Our analysis is a case study in augmenting economic models to incorporate news attention and retrieve narratives. But these ideas apply to other modeling approaches as well. For example, the implied residuals from an estimated DSGE model can be projected onto news attention to extract narratives that aid interpretation of those shocks. More generally, by combining macroeconomic analysis with topic modeling, the researcher can draw on vast written text corpora to better understand quantitative economic phenomena. Our results provide a glimpse into the possibilities of using textual data for modeling macroeconomic dynamics or the role of information transmission and the media in the macroeconomy. To help facilitate such work, we maintain the interactive website, www.structureofnews.com , to allow users to visualize and inspect a wide variety of features from our estimated topic model. In addition, the website allows researchers to download our *WSJ* news attention time series (daily and monthly) for use in their own projects.

Our work contributes to a rapidly growing literature in economics that uses text as data.[3] Topic models have only recently begun to be explored in empirical economics research (the earliest example to our knowledge is Hansen, McMahon, and Prat (2017)). Papers by Larsen and Thorsrud (2019) and Thorsrud (2020) apply LDA to Norwegian news data and analyze macroeconomic forecasting models.[4] Complementary research integrates news text into macrofinance analyses using carefully curated researcher inputs in place of statistical models. Chahrour, Nimark, and Pitschner (2021) study the influence of newspapers' sectoral coverage on macroeconomic activity. Rather than directly studying article text, they map article tags from Factiva to firms and their respective sectors in order to measure news coverage. Baker, Bloom, and Davis (2016) build indices of EPU from news articles by counting the occurrences of researcher-curated key words, and Baker et al. (2021) analyze the drivers of large stock price moves based on close human reading of business news. In this paper, we provide a complete solution for macroeconomic textual analysis to better understand the forces driving business cycles.[5] This includes (i) estimating and selecting an optimal topic model specification from a macroeconomic news corpus, (ii) a text-augmented VAR that embeds the topic structure in an otherwise standard empirical macroeconomic model, (iii) a selection algorithm for deciding the appropriate topics to include in the VAR, and (iv) a means of extracting underlying articles that provide narrative interpretations of business cycle fluctuations.

We also contribute to a strand of literature that uses text data to forecast stock market returns. García (2013) shows that news sentiment forecasts stock market returns especially well during recessions. Manela and Moreira (2017) show that news-implied volatility forecasts returns and economic disasters. Ke, Kelly, and Xiu (2019) and Kelly, Manela, and Moreira (2021) propose novel machine learning methodologies that convert raw news text into forecasts for stock returns and macroeconomic indicators, respectively. Boudoukh et al. (2018) explain contemporaneous idiosyncratic returns for individual stocks using news text. We contribute to this literature by showing that topic models distill reliable market-timing signals from news text. We analyze "online" methods for forecasting with topic models while avoiding look-ahead bias, and for interpreting market-timing gains through the lens of economically interpretable narratives.

Section I briefly discusses our model structure and estimation. Section II describes the estimated structure of economic news underlying the *WSJ* text.

[3] See Gentzkow, Kelly, and Taddy (2019) for a recent survey.

[4] Subsequent to our paper, Ellingsen, Larsen, and Thorsrud (2020) also analyze the *WSJ* news for macroeconomic forecasting.

[5] While we use LDA and analyze the full text of all *WSJ* articles, Cong, Liang, and Zhang (2020) use word2vec methods and Manela and Moreira (2017) use support vector machines, with both of these papers analyzing abstracts of front-page articles.

Section III documents the close correspondence of news text and numerical economic data. Section IV analyzes the role of news text in a macroeconomic VAR. Section V analyzes market timing with news text. Section VI concludes.

I. Topic Model and Estimation

Our model for the structure of economic news text follows the LDA topic modeling approach of Blei, Ng, and Jordan (2003). We provide a brief summary of the model and its estimation, and refer interested readers to the original paper for additional details.

We study the *WSJ* text corpus in its "bag-of-words" form, represented numerically as the article-term matrix, $\boldsymbol{w}$. This is a $T \times V$ matrix, where row indices correspond to the T different articles in the corpus, and column indices correspond to the V unique terms in the corpus. Its individual elements, $w_{t,v}$, count the number of times that the v^{th} term appears in article t.

A. LDA

While the bag-of-words representation is a dramatic reduction in complexity compared to the raw text, it remains an extraordinarily high-dimensional object. LDA seeks a tractable thematic summary of $\boldsymbol{w}$ that reduces the dimensionality of the historical *WSJ* to a scale that can be digested and interpreted by a human reader in one sitting. To achieve this summary, it imposes explicit parametric assumptions (a multinomial-term-count distribution with Dirichlet priors) and imposes a factor structure on expected term counts. Specifically, LDA assumes that the V-dimensional vector of term counts for a given article t, denoted w_t, is distributed according to a multinomial distribution

$$w_t \sim \text{Mult}(\Phi'\theta_t, N_t), \tag{1}$$

where N_t is the total number of terms in article t and governs the scale of the multinomial distribution. In other words, expected term counts are summarized by a comparatively low-dimension set of parameters, θ_t and $\boldsymbol{\Phi} = [\phi_1, \ldots, \phi_K]'$. The k^{th} "topic" in the text is defined by the V-dimensional parameter vector ϕ_k, where $\phi_{k,v} \geq 0$ for all v and $\sum_v \phi_{k,v} = 1$. That is, a news topic is a probability distribution over terms. The set of terms that take especially high probabilities in ϕ_k convey the thematic content of the topic. The model's dimension reduction is achieved by setting the total number of topics K much smaller than the size of the vocabulary.

While the topics, ϕ_k, describe the common themes in the corpus as a whole, LDA treats an individual article as a mixture of topics. The article-specific parameter vector $\theta_t = (\theta_{t,1}, \ldots, \theta_{t,K})'$ is also a probability vector, with $\theta_{t,k} \geq 0$ for all k and $\sum_k \theta_{t,k} = 1$. That is, θ_t describes how article t allocates its attention across topics. LDA embeds a factor structure in which topics (ϕ_k) serve as common factors and θ_t captures article-specific exposure to those factors.

In principle, one could estimate ϕ and θ via maximum likelihood according to (1). However, in most text applications this is computationally unrealistic. Instead, we use Bayesian methods and approach estimation with the collapsed Gibbs sampler proposed by Xiao and Stibor (2010). We describe this approach and how it generates estimates of the topics, $\hat{\phi}_k$, and estimates of attention allocation across topics, $\hat{\theta}_t$, in Appendix IA.B.

B. The WSJ *Data Set*

The data set we use is among the most extensive text corpora of business news studied in the economics literature to date. It consists of all articles published in the *WSJ* from January 1984 through June 2017, purchased from the Dow Jones Historical News Archive. This represents the longest history of full text articles in digital text format available for purchase from Dow Jones & Company. As a comparison, Manela and Moreira (2017) extract a news-based analog to the VIX index using the *WSJ* over a long sample, but their data only include abstracts of front-page articles. Baker, Bloom, and Davis (2016) study a large collection of newspapers, but restrict their analysis to counting occurrences of a small, predefined key term list. In contrast, our analysis leverages the richness of full newspaper text.

We take a number of steps to homogenize the data sample and reduce confounding effects of organizational changes to the *WSJ* over time. First, the Dow Jones Historical News Archive contains data as far back as 1979, but data prior to 1984 are limited to article abstracts. We therefore omit pre-1984 data to maintain consistency in the definition of what constitutes an article throughout our sample. Next, over its history, *WSJ* has initiated (and sometimes later abandoned) a number of noncore sections such as "Personal Journal" (initiated 2002), "Weekend Journal" (initiated 2005), and "Off Duty" (initiated 2012). To help maintain consistency in topical content over time, we exclude articles appearing in sections other than the three core sections ("Section One," "Marketplace," and "Money and Investing") that are available over our full sample. Because our interest is in economic news, we also exclude articles with subject tags corresponding to predominantly noneconomic content such as sports, leisure, and arts, as well as articles that contain data tables with little supporting text.

Next, we transform the data set from a collection of raw article text files into numerical term counts for statistical modeling. Our vocabulary of "terms" includes all unigrams and bigrams occurring in our data set after applying a mild set of term filters and lemmatization of derivative words. Appendix IA.A provides a step-by-step description of the data-processing procedure we use to transform raw article text into term counts.

Our final data set consists of 763,887 articles with a vocabulary of 18,432 unique terms. The left panel of Figure 1 shows the size of the data set over time in terms of monthly article count and monthly term count. The structure of the topic model absorbs some of the secular changes in news production by modeling news term proportions rather than term count levels. It is worth

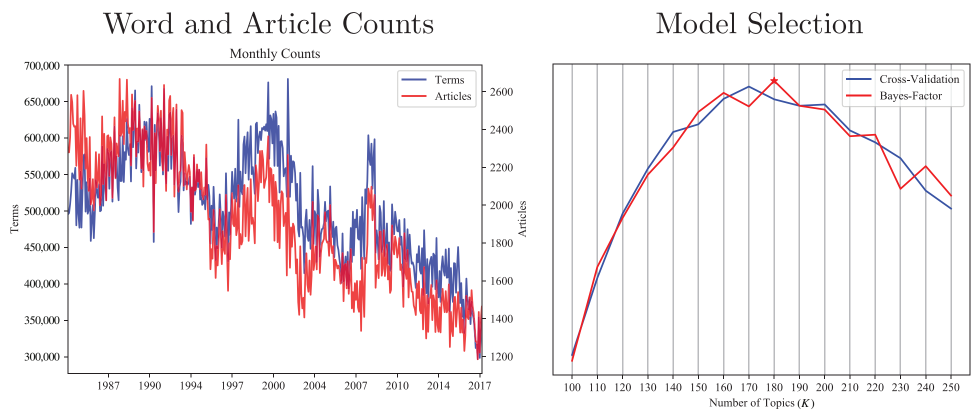


Figure 1. *WSJ* observation counts and model selection. The left panel plots the postprocessing article count and the total term counts aggregated over all articles each month. The right panel plots the cross-validated model fit (average log-likelihood over 10 cross-validation folds) and the Bayes factor for models with topic count K ranging from 50 to 250 in increments of 10 (we omit labels on the right panel vertical axis as the model fit comparisons are relative and specific values are uninformative). (Color figure can be viewed at wileyonlinelibrary.com)

noting that, our homogenization efforts notwithstanding, the *WSJ* is an evolving product that undergoes structural shifts over the course of our sample.

C. Topic Model Specification Choice

LDA is an unsupervised machine learning technique that requires the econometrician to choose only two inputs: the term count data set and the number of topics, K. We take a data-driven approach to select K. In particular, we estimate a variety of models with K ranging from 50 to 250 topics in increments of 10. We select the model specification from this set as the one with the highest Bayes factor.[6] For robustness, we also approach model selection with 10-fold cross-validation. In particular, we partition the *WSJ* sample into 10 folds with equal numbers of articles. We reestimate each K-topic model on the data 10 times, one time excluding each of the 10 validation samples, then evaluate its fit on the left-out sample. We calculate aggregate goodness-of-fit for each model as the average log-likelihood value over the 10 validation samples.

The right panel of Figure 1 plots goodness-of-fit across candidate specifications. Bayes factor and cross-validation analyses indicate that a 170- to 180-topic specification approximately optimizes our specification criteria. Manual inspection of estimates for various choices of K also indicates that

[6] The Bayes factor is the ratio of posterior probabilities for the alternative model versus the null model. When selecting among topic models with different K, we compare all models to the same null model so that the denominator is constant across K. Thus, our Bayes factor criterion is equivalent to selecting the model with the highest posterior probability.

$K = 180$ is a sensible choice. A model with fewer topics (e.g., 50) produces mixed topics that contain multiple separate themes. A larger model (e.g., 250 topics) delivers similar interpretability as that for the 180-topic model, but with a number of overly specific topics that capture one-off events. Finally, all models in the neighborhood of $K = 180$ look very similar in their topic composition. Appendix IA.B provides further visual evidence that the 180-topic model strikes a balance between interpretability and parsimony.

D. oLDA for Avoiding Look-Ahead Bias

In many analyses common in finance and economics, researchers wish to quantify the degree to which variables observable at a point in time can forecast future outcomes. A canonical example is market timing—forecasting future stock market returns using only information available in a prior period. For example, below we study one such application, and find that topic attention observed in month t can forecast stock market returns at $t + 1$.

A common concern in forecasting analysis is look-ahead bias. If the variables used for prediction are constructed using information available only in future periods, then forecasting regression estimates could be biased. In the LDA topic model context, if we use the entire sample to estimate the topics $\boldsymbol{\Phi}$ and attention allocation across topics θ_t, then $\hat{\theta}_t$ is informed in part by term counts that realize after time t.[7]

To overcome this bias, we turn to an oLDA estimator from Hoffman, Bach, and Blei (2010) that constructs topic model estimates in an entirely backward-looking fashion. While oLDA was originally developed to overcome computational challenges associated with data sets too large to fit in memory, we take advantage of its ability to process text data that arrive sequentially, and emphasize this advantage for out-of-sample forecasting analysis.

The key benefit of oLDA for our purposes is that it processes documents in sequential batches. Therefore, time t topic attention estimates are based only on data from periods 1 through t. As a result, time t topic attention is a valid input for forming out-of-sample forecasts of $t + 1$ outcomes without look-ahead bias.[8] Moreover, unlike rolling reestimation, which faces the issue that topics in one rolling sample are delinked from topics in a prior sample, oLDA preserves a coherent topic interpretation over time (while accommodating evolving topics). We refer readers to Hoffman, Bach, and Blei (2010) for additional details.

[7] A standard finance analogy is using PCA to estimate return factors from a matrix R of N stock returns over T periods. The factor estimates for any date $t < T$ are indirectly informed by data after date t because the PCA factor loadings use information from the full sample.

[8] To estimate oLDA, we use the *gensim* python package, which employs an alternative estimation procedure—variational inference as opposed to Gibbs sampling. However, manual inspection of the resulting topics suggests that both estimation procedures yield comparable topic attention.

II. The Structure of Economic News

In this section, we dissect the *WSJ* news topic model estimates. We focus here on the 180-topic specification based on a single estimation using the full sample, and discuss oLDA estimates in Sections IV and V. Note that unlike other dimension-reduction techniques such as PCA, LDA does not provide a natural ordering of topics, and thus our presentation of topics is based on expositional convenience. It is also important to recognize that due to the richness of information contained in news text, our 180-topic model is higher-dimensional than typical economic models, and it is impossible to report the full scope of estimates in this paper. To give readers the ability to explore all facets of our model, we have provided an interactive website, www.structureofnews.com , to allow users to inspect our estimated topic model in detail.

A. Topic Key Terms

First, we present estimates for each term cluster—that is, the $V \times 1$ term probability vector ϕ_k—that defines a given topic k. The most common terms in the vocabulary appear with high frequency in many topics, and this is naturally reflected in LDA estimates. To best identify the *unique* semantic content of each topic, we rescale the topic-term weights by the inverse of the term frequency, f_v:

$$\tilde{\phi}_{k,v} = \frac{\hat{\phi}_{k,v}}{f_v}.$$

This scaling emphasizes terms that have an unusually large weight in topic k (terms like "pharmaceutical" or "iron ore") and downplays words that are common to the corpus overall (terms like "price" and "company"). Sorting the elements of $\tilde{\phi}_k$ identifies the terms that are most diagnostic of the thematic content in topic k.

Internet Appendix Table IA.I lists the *key terms* for each topic k, defined as the top 10 vocabulary items based on the sorted $\tilde{\phi}_k$ vector. These key terms are entirely unsupervised as they are estimated from the data without any guidance from article labels or from the researcher. We manually assign a label to each topic based on our reading of the key term lists. Topic labels serve as a shorthand for referencing topics throughout the paper.

Several features of the key term lists stand out. First, they show that topics represent coherent concepts with clear interpretability. For example, the first key term list from the table is

> *bonus, base salary, total compensation, pay package, compensation package, compensation committee, restrict stock, tyco, executive compensation, bonuses*.

This list is easily recognized as an "executive pay" topic. The list is not mixed. Each term has a direct narrative link to the topic label. The purity of the topic extends beyond the top 10 terms. For example, key terms 11–25 for this topic are

stockoption, executive pay, stock option, compensation, annual salary, tyco international, exercise option, severance package, severance payment, exercise price, salary, retention, vest, proxy statement, severance.

This term purity is representative of our estimated model more broadly. No topics appear to mix themes. To draw a contrast, when we estimate a smaller model with $K = 50$ topics, we find that topics begin to lump multiple concepts together. For example, the 50-topic model generates the following key term list for one of its topics:

creditor lawsuit, unsecured creditor, ual corps, chapter company, texas air, bankruptcy court, federal bankruptcy, bankruptcy code, continental airline, usair, load factor, flight attendant, reorganization plan, american airline, northwest airline, chapter file, amr corps, bankruptcy protection, major airline, corps unit, airline unit, pilot association, chapter bankruptcy, world airways, twa.

Evidently, this topic combines two distinct topics—"airlines" and "bankruptcy"—that we identify separately in the $K = 180$ specification. While this mixing is understandable given the string of airline bankruptcies in our sample period, mixing confounds the model's ability to identify important nonbankruptcy airline news such as sharp drops in flight demand following terrorist events. Examples like this abound in smaller models.[9] The general absence of mixing in the $K = 180$ specification supports the view that the larger model achieves a successful separation of distinct subjects.

Second, we see that most topics represent news subjects, as opposed to appraisals of whether news for a given subject is good or bad. Examples of appraisal-free subject topics include "airlines," "Federal Reserve," and "China." Some subject topics carry an implicit appraisal, usually of bad news, such as "terrorism," "natural disasters," and "recession." A small number of sentiment-related topics also appear, such as the "problems," "concerns," and "positive sentiment" topics. We interpret these as "modifier" topics that, when combined in a news article with a given subject topic, convey a signed or directional appraisal of the subject's current events.[10]

[9] Another example in the $K = 50$ specification is a topic with key terms

ipo market, sachs group, ipo, csfb, stearns cos, berkshire, buffett, berkshire hathaway, suisse group, ipo price, brother hold, investmentbanking business, stanley group, goldman, hathaway, group credit, firstday, weill, witter discover, investmentbanking, stanley dean, goldman sachs, sachs, warren buffett, credit suisse,

which is a mixture of the "IPO," "investment banking," and "Buffett" topics in the $K = 180$ model.

[10] A very small number of topics that represent shifts in language usage over time, but do not have any economic meaning. For example, there is a "corrections/amplifications" topic {key terms: *article correction, amplifications, incorrect, misstated*} and a "news conference" topic {key terms: *press conference, told reporter, apology, issue statement*}. Both of these have a clear positive time

A larger K results in more specific subject categories. An advantage of having fine-grained topics is that, if a broader notion of a topic is desired, finer topics can be combined into broader metatopics. Indeed, we find that our 180 topics cluster into an intuitive hierarchy of increasingly broad metatopics. Figure 2 illustrates the topical hierarchy in our model in the form of a dendrogram. We estimate this hierarchy with recursive agglomeration (Murtagh and Legendre (2014)) based on the semantic distance between topics, defined as the distance in their $\hat{\phi}_k$ vectors. Like the estimation of topics themselves, the connectivity of topics in the dendrogram is entirely data-driven, and we only supply metatopic labels based on our manual read of the clusters. While this figure focuses on the high-level metatopic hierarchy, interested readers can find the metatopic assignment of all 180 individual topics in Table IA.I.

Figure 2 provides a full taxonomy of the *WSJ* news. At the broadest level, news is classified into either "economy" topics (top half of the dendrogram) or "politics and culture" topics (bottom half). Within "economy," topics split into broad metatopics such as "financial intermediaries," "economic growth," and "industry." The "politics and culture" branch includes topic clusters such as "international relations," "national politics," and "science and arts." As emphasized by Quinn et al. (2010), an intuitive metatopic hierarchy is a useful check on the semantic validity of a model. Furthermore, it gives us the ability to analyze attention to news topics at various levels of granularity.

B. Quantifying News Attention

With an understanding of the topic structure in place, we turn to the second main output of the model—the estimate of news attention paid to each topic. Attention allocation is a convenient quantitative transformation of news composition that can serve as numerical input to subsequent empirical investigations of economic hypotheses. Our estimates describe how allocation of media attention across topics evolves over time.

B.1. News Attention by Month

Topic attention is estimated at the article level but can be summed to any level of aggregation according to equation (IA.3). We focus on aggregate *WSJ* topic attention at the monthly frequency by summing attention estimates over all articles published in the same calendar month. Figure 3 shows the time-series variation in attention for a subset of six illustrative topics.[11] The black line, corresponding to the left vertical axis, shows the attention to a given topic as a percent of total monthly *WSJ* news production. The right

trend. Language topics are valuable for denoising news text by absorbing common variation in language that is distinct from genuine news content and that would otherwise be absorbed into a content topic and act as a source of noise in our subsequent analysis.

[11] Figure IA.1 plots an additional 24 illustrative topic attention time series. The website for this research project, www.structureofnews.com, includes time-series plots for all 180 topics.

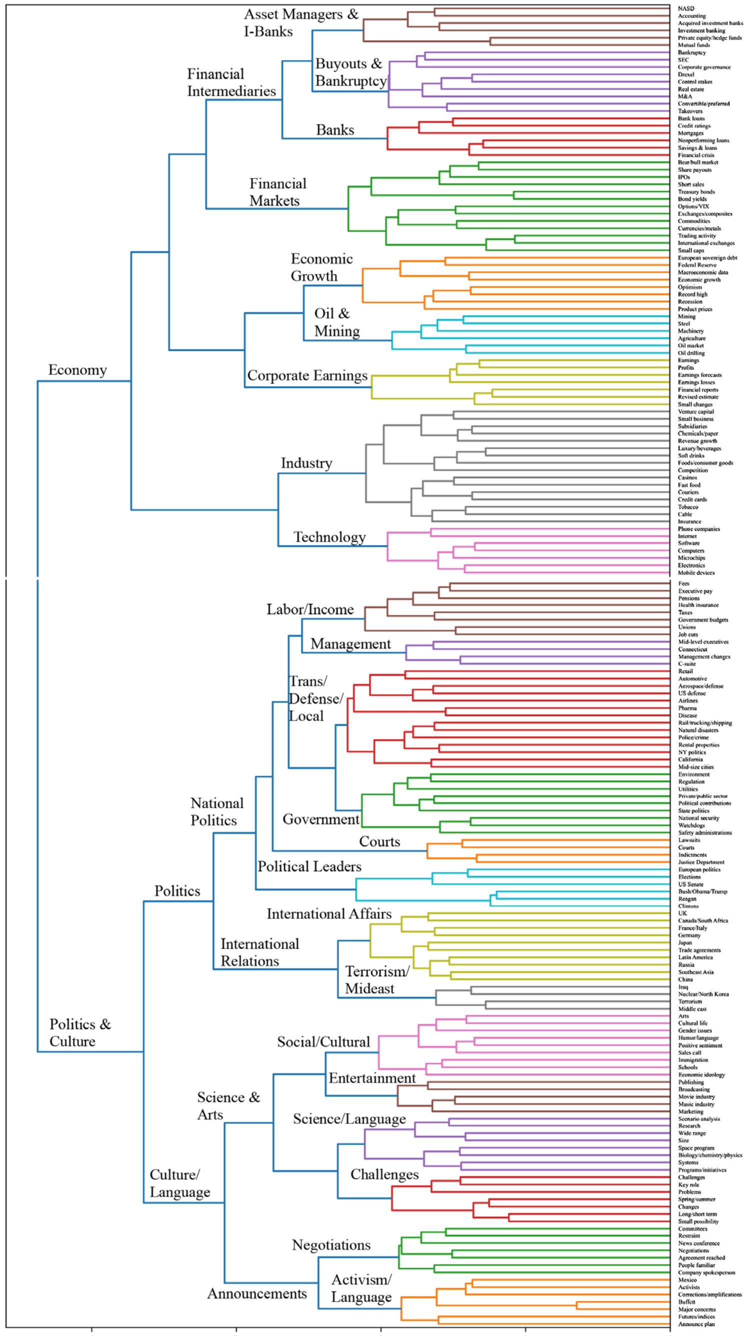


Figure 2. Hierarchical taxonomy of the *WSJ* news. Hierarchical agglomerative clustering dendrogram based on $\hat{\phi}_k$ similarity among the 180 topics listed on the right. (Color figure can be viewed at wileyonlinelibrary.com)

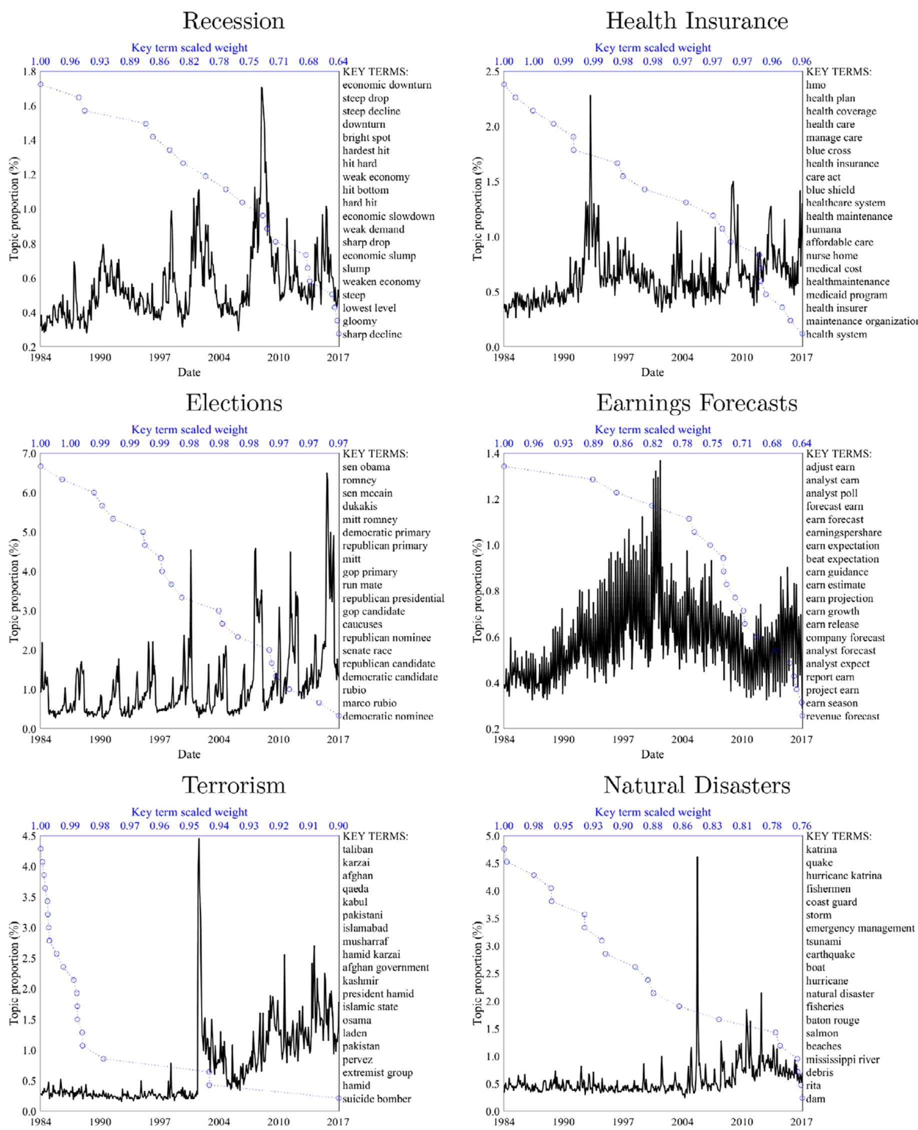


Figure 3. Topic attention proportions. The black line shows topic attention as a percent of total monthly *WSJ* news production. The right vertical axis lists the topic's key terms, and the blue line shows the weight of each term in the topic (scaled by the maximum term weight within that topic). (Color figure can be viewed at wileyonlinelibrary.com)

vertical axis lists the key terms, and the blue line shows the weight of each term in the topic (scaled by the maximum term weight within that topic), corresponding to the top horizontal axis.

The topic attention plots highlight a number of stylized facts about the composition of business news. First, news attention is generally persistent. This is clearly evident in the top two topics, "recession" and "health insurance," which exhibit prolonged waves of high and low attention. "Recession" is a prime example of a recurrent topic that is active throughout the sample, as is the "health insurance" topic, which becomes highly prevalent during debate of the Clinton Health Plan proposal (peaking around President Clinton's speech to Congress in September 1993), the proposal and passage of Obamacare in 2008 to 2010, and around the debated repealing of Obamacare around the 2016 presidential election. The middle two plots in Figure 3 are examples of seasonal topics. "Elections" coverage has a cyclical pattern peaking every four years with a secondary peak every two years, and the "earnings forecasts" topic spikes ahead of each quarterly earnings announcement season. The last two plots are examples of emergent topics. Usually dormant, they are punctuated by intense coverage around particular events. The "terrorism" topic is a prototypical regime shift: it receives minute attention in the first half of our sample, spikes dramatically on 9/11, and remains high thereafter. Similarly, the "natural disasters" topic draws little attention during most of the sample, but rises sharply in August 2005 (Hurricane Katrina), before reverting but remaining slightly elevated for much of the remaining sample, presumably due to increased news attention on climate change (Engle et al. (2020)).

With 180 topics, it is difficult to isolate the focus of attention at each point in time in a single line plot overlaying many time series. To make these news attention series easier to grasp, in Internet Appendix IA.D we report the dominant topics in each month and in Internet Appendix IA.E we plot attention at the metatopic level.

III. Matching News Attention to Economic Activity

We next validate our topic model estimates by investigating how faithfully the topical content of the *WSJ* news captures the state of the economy. We present a descriptive analysis documenting the high correlation between news text and a wide range of numerical economic time series.

A. Selection via Lasso Regression

As a first step toward establishing that news attention beneficially measures the state of the economy, we document that diverse numerical measurements of economic activity all share a close correspondence with thematically related topical content of news text.

To this end, we devise a generic means of evaluating the alignment between a candidate numerical measure of economic activity, x_t, and the set of estimated topic attention proportions, $\hat{\theta}_t = (\hat{\theta}_{1,t}, \ldots, \hat{\theta}_{K,t})$. Our criteria for this analysis are twofold. First, we evaluate accuracy of the text-based explanation in terms of regression fit, and second, we evaluate whether the most influential news topics in the regression are thematically related to x_t.

With $K = 180$ topics, the attention vector is high-dimensional relative to the number of monthly time-series observations, so ordinary least squares regression of x_t on $\hat{\theta}_t$ will be difficult to interpret and suffer from overfit. We therefore use lasso regression and set the lasso penalty parameter such that exactly five of the 180 coefficients take nonzero values. We choose a fixed five-variable regression to ensure clear interpretability of results, to have uniformity and comparability across the different x variables that we study, and to limit the undue influence of data mining on our conclusions.

Because lasso penalization serves in part as a model selection device, standard errors on the nonzero coefficients must be adjusted to account for model search. Our reported p-values use the postselection inference procedure of Tibshirani et al. (2016). We scale regression coefficients to reflect the dependent variable's response in standard-deviation units to a one-standard-deviation change in the regressor.

To assess the breadth of economic summary achieved by studying news text, we examine a diverse set of numerical economic data series. The gamut of variables that we study includes macroeconomic aggregates, corporate financing activities, measures of industry-level risks, and uncertainty about a range of economic policies.

B. Macroeconomic Aggregates

We begin by studying monthly data on aggregate output, employment, stock market returns, and stock market volatility. The top left of Table I reports results for the lasso regression of log industrial production growth on the 180 news attention time series. Among the five selected topics, news attention to the "recession" and "oil market" topics significantly negatively correlates with output growth. The coefficient estimates are economically large. A one-standard-deviation increase in "recession" attention associates with a 0.38 standard deviation drop in output growth.

In fact, "recession" attention is by far the strongest explanatory variable for all four economic aggregates in Table I. A one-standard-deviation rise in "recession" attention associates with a 0.61 standard deviation drop in log employment (nonfarm payrolls) growth. Employment growth also has a significant positive coefficient on "rail/trucking/shipping" attention and a negative coefficient on "Iraq" attention, although the impact of these topics is small compared to that of "recession" attention.[12]

The lower left panel regresses returns on the value-weighted U.S. stock market index on AR(1) innovations in topic attention.[13] "Recession" attention

[12] The lasso regression juxtaposes the employment experience under various presidential administrations, although these effects are insignificant at the 5% level. This result also shows how the choice of topic number can influence second-stage economic analysis. A finer decomposition of the *WSJ* into 300 topics produces separate topics for each presidential administration, while a coarser decomposition of 50 topics groups all administrations together.

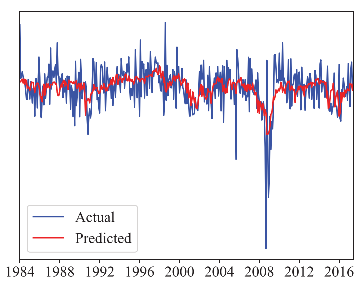
[13] Market returns have close to zero serial correlation and theoretically behave like a martingale difference sequence, making it natural to likewise represent the regressors as time-series innovations.

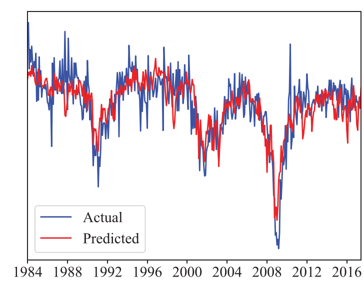
Table I

Reconstructing Macroeconomic Time Series

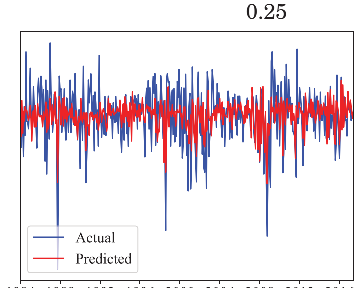
Five-regressor active-set regression estimates based on lasso selection with p-values adjusted for postselection inference.

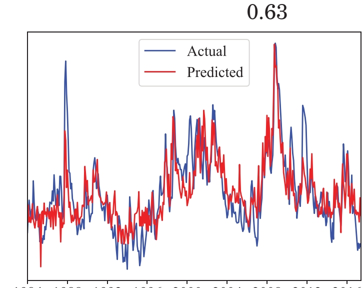
Industrial Production Growth			Employment Growth		
Topic	Coeff.	p-Val.	Topic	Coeff.	p-Val.
Recession	−0.38	0.00	Recession	−0.61	0.00
Oil market	−0.17	0.00	Rail/trucking/shipping	0.22	0.01
Southeast Asia	0.11	0.10	Bush/Obama/Trump	−0.15	0.09
Health insurance	0.06	0.93	Iraq	−0.14	0.01
Clintons	0.03	0.40	Clintons	0.12	0.01
R^2	0.21		R^2	0.59	





Market Returns			Market Volatility		
Topic	Coeff.	p-Val.	Topic	Coeff.	p-Val.
Recession	−0.35	0.00	Recession	0.48	0.00
Problems	−0.19	0.00	Options/VIX	0.24	0.00
Convertible/preferred	−0.16	0.21	Problems	0.22	0.00
Record high	0.12	0.08	Small business	−0.18	0.00
Options/VIX	−0.11	0.76	Investment banking	0.13	0.00
R^2	0.25		R^2	0.63	





is the largest determinant of market returns with a coefficient of −0.35. The second-largest effect comes through attention to the "problems" topic, which describes adverse conditions more generally {key terms: *problem face, biggest problem, big problem, matter worse*}.

Table II

Reconstructing Stock Market Returns

Five-regressor active-set regression estimates based on lasso selection with p-values adjusted for postselection inference. Full R^2 corresponds to the fit with both topics and benchmark variables. Benchmark R^2 corresponds to the fit with only benchmark variables.

Topic Model & CFNAI			Topic Model & FRED-MD		
Topic	Coeff.	p-Val.	Topic	Coeff.	p-Val.
Recession	−0.35	0.00	Recession	−0.36	0.00
Problems	−0.19	0.00	Problems	−0.18	0.00
Convertible/preferred	−0.16	0.20	Convertible/preferred	−0.14	0.24
Record high	0.12	0.08	Record high	0.13	0.07
Options/VIX	−0.11	0.76	IPNCONGD (FRED-MD)	−0.10	0.98
Full R^2	0.25		Full R^2	0.25	
Benchmark R^2	0.02		Benchmark R^2	0.09	

"Recession" and "problems" attention are also significant positive drivers of market volatility, defined as the realized standard deviation of daily market returns within each month. They both enter with large positive coefficients, reflecting the strongly countercyclical nature of market volatility. It also shows that "small business" topic tends to associate with low volatility regimes in our sample.

For all four macroeconomic variables, five news topics explain a significant fraction of their variation measured by R^2. News attention explains 25% of the variation in stock market fluctuations. To illustrate, Table II compares news attention to other business cycle leading indicators. We first regress stock returns on AR(1) innovations in the four subindices of the Chicago Fed National Activity Index (CFNAI). These indices measure production/income, employment/unemployment/hours, personal consumption/housing, and sales/orders/inventory. CFNAI indices jointly explain only 2% of the variation in contemporaneous monthly stock returns. Table II also shows that when we include these along with the news attention series in our five-regressor lasso model, none of the CFNAI indices are selected.

As another benchmark, we regress market returns on AR(1) innovations in the 101 macroeconomic variables in the FRED-MD database.[14] To remain comparable with our news attention regressions, we likewise impose a lasso penalty to select exactly five nonzero coefficients among the FRED-MD variables. We find an R^2 of 9%, which exceeds that from the CFNAI regression but is less than half of the news-based result.[15] When we include the 101

[14] See https://research.stlouisfed.org/econ/mccracken/fred-databases/. We exclude FRED-MD interest rate and exchange rate variables as they are asset prices and have mechanical links to stock market returns.

[15] This result is robust to a variety of regression specifications. The explained variation in returns is essentially unchanged if we use other constructions of macro variable innovations, or if we control for lagged as well as contemporaneous values. Even if we include interest rate and exchange rate variables, the R^2 rises only to 17%.

FRED-MD series alongside the news attention series in the lasso model, only one FRED-MD variable is selected, namely, industrial production growth of nondurable consumption goods (IPNCONGD), though it is insignificant and enters with a counterintuitive negative sign.

C. Financing Activity

As a second and more specific example, we analyze how news attention tracks fluctuations in financing activity. We choose x_t variables that help illustrate how specific news topics line up with detailed differences across economic activities.

First, we study monthly dollar volumes of IPOs and LBOs. Table III shows that news attention closely tracks fluctuations in both series with an R^2 of 44% and 58%, respectively. More interestingly, the news topics that associate most strongly with each financing variable are also those with the most similar thematic content. The three significant topics in the IPO volume regression are "IPOs," "venture capital," which is the typical form of financing prior to IPO, and "internet," which is the industry that saw the greatest IPO activity during our sample period. For LBO volume, the "takeovers," "control stakes," "key role," and "job cuts" topics capture key LBO concepts such as transition from public market to private ownership and control, management changes ("key role"), and eliminating jobs to reduce costs and improve production efficiency.

Table III also looks at the credit risk side of financing, measured first as the monthly count of bankruptcies among U.S. public companies. News attention explains variation in bankruptcy intensity with an R^2 of 42%, and the four significantly associated topics ("recession," "accounting," "venture capital," and "small caps") are all themes that commonly coincide with debt default. The last panel introduces an international dimension to credit risk in the form of average credit default swap (CDS) spreads among Eurozone sovereigns (available beginning in 2002). A one-standard-deviation rise in the lone significant news topic, "European sovereign debt," associates with a 0.58 standard deviation increase in European sovereign credit spreads. The time-series plot of actual spreads versus spreads reconstructed from news shows the rapid rise and fall of CDS spreads. This data series amounts to only a handful of distinct observations around the financial crisis. Notwithstanding, estimated news topic attention is informative enough that lasso regression hones in on the correct narrative account of these events. While other news topics related to financial crises and subprime mortgages follow similar time-series patterns, the penalized regression bypasses these in favor of the European sovereign debt narrative.

D. Industry Volatility

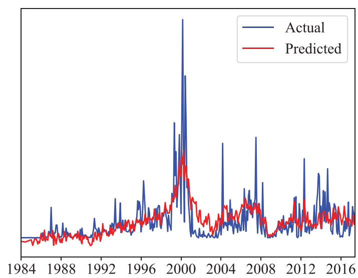
Our next example investigates how well news attention accounts for patterns in industry-level stock volatility. We use the 49-industry categories

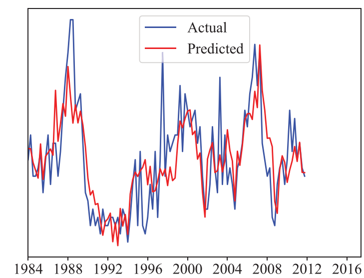
Table III

Reconstructing Financial Activity

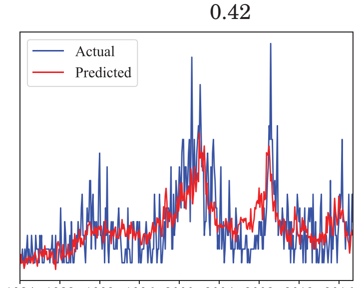
Five-regressor active-set regression estimates based on lasso selection with *p*-values adjusted for postselection inference.

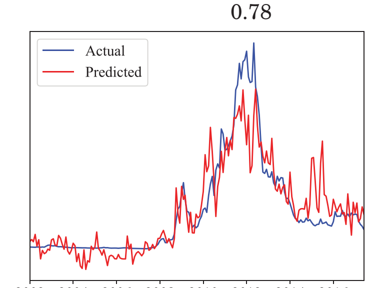
Topic	Coeff.	*p*-Val.	Topic	Coeff.	*p*-Val.
IPO Volume			LBO Volume		
IPOs	0.26	0.00	Takeovers	0.31	0.28
Venture cap.	0.22	0.00	Insurance	−0.28	0.02
Bankruptcy	−0.14	0.25	Control stakes	0.27	0.03
Internet	0.13	0.08	Key role	0.23	0.70
M&A	0.09	0.72	Job cuts	−0.18	0.13
R^2	0.44		R^2	0.58	





Topic	Coeff.	*p*-Val.	Topic	Coeff.	*p*-Val.
Bankruptcy Filings			European CDS Spreads		
Recession	0.48	0.00	Euro. sov. debt	0.58	0.00
Venture cap.	0.25	0.01	Govt. budgets	0.17	0.61
Accounting	0.24	0.01	Middle east	0.16	0.08
Small caps	0.12	0.00	Mobile devices	0.09	0.06
Machinery	−0.06	0.31	News conference	0.07	0.98
R^2	0.42		R^2	0.78	





from Ken French's website, and measure monthly volatility as the standard deviation of daily industry returns within the month. All equity portfolios tend to share a large common time-series component associated with overall market volatility (Herskovic et al. (2016)). To hone in on industry-specific volatility patterns that are distinct from the aggregate market volatility analysis in Table I, we perform two adjustments to the raw industry volatility

Table IV

Reconstructing Industry Volatility (Innovations)

Five-regressor active-set regression estimates based on lasso selection with p-values adjusted for postselection inference.

Topic	Coeff.	p-Val.	Topic	Coeff.	p-Val.	Topic	Coeff.	p-Val.
Automotive			Banking			Pharmaceuticals		
Automotive	0.16	0.03	Nonperform. loans	0.26	0.00	Clintons	0.20	0.01
Fees	0.14	0.11	Mortgages	0.20	0.00	News conf.	−0.14	0.48
Mutual funds	−0.13	0.33	Options/VIX	−0.18	0.00	Pharma	0.14	0.10
Corporate govt.	0.12	0.59	Chemicals/paper	−0.15	0.10	Earnings fcts.	0.12	0.35
Financial crisis	−0.08	0.87	NASD	−0.09	0.76	Health ins.	0.10	0.62
R^2	0.10		R^2	0.17		R^2	0.11	
Computer Hardware			Oil and Gas			Tobacco		
Retail	0.20	0.00	Oil market	0.34	0.00	Tobacco	0.17	0.34
Computers	0.17	0.01	Earnings losses	−0.15	0.05	Problems	−0.10	0.43
Software	0.10	0.18	Options/VIX	−0.15	0.87	Mexico	0.10	0.98
Fees	−0.09	0.72	Cultural life	0.14	0.08	Terrorism	−0.09	0.41
Russia	−0.08	0.93	Elections	−0.13	0.46	Lawsuits	0.09	0.59
R^2	0.09		R^2	0.21		R^2	0.08	

data. First, we orthogonalize each industry volatility series against the first principal component of the industry volatility panel. Second, we construct the AR(1) innovations in the adjusted series.[16]

Table IV reports results from five-regressor lasso regressions of industry volatility innovations on AR(1) innovations in news topic attention. For the sake of brevity, the table reports results for nine of the 49 industries that best illustrate how specific aspects of news attention align with specific industry behaviors. We find that the influential topics in each regression are indeed close thematic counterparts for the respective industry. For example, banking sector volatility is highest amid news attention to the "nonperforming loans" and "mortgages" topics. Volatility in the computer hardware industry (which includes firms like Apple, Dell, and Hewlett Packard) is most tightly linked to "retail," "computers," and "software" topic attention. In addition to the "pharma" news topic, volatility in the pharmaceuticals sector is associated with attention to the "Clintons" topic, reflecting the important role of health care reform in the policy platforms of both Bill and Hillary Clinton, as well as attention to the "health insurance" topic. Occasionally, we find unintuitive but significant topic assignments such as "cultural life" in the oil and gas industry regression, although these examples are rare.

[16] Without the principal component adjustment, we essentially recover the results seen for market volatility in Table I. While the principal component adjustment is necessary for our conclusions, the results from our analysis are largely unchanged if we use levels of orthogonalized industry volatility rather than AR(1) innovations.

Table V

Reconstructing Macroeconomic Time Series with Interactions

Five-regressor lasso regression estimates and fits with p-values are adjusted for postselection inference, using individual topic attention regressors as well as their pairwise interactions.

Industrial Production Growth			Employment Growth		
Topic	Coeff.	p-Val.	Topic	Coeff.	p-Val.
Recession × Oil market	−0.25	0.00	Inv. banks × Bush/Obama/Trump	−0.28	0.94
Inv. banks × Bush/Obama/Trump	−0.18	0.17	Recession × Fees	−0.22	0.24
Job cuts × Recession	−0.09	0.73	Publishing × Recession	−0.19	0.51
Recession × Corporate governance	−0.07	0.94	Recession × Bankruptcy	−0.17	0.04
Short sales × Recession	−0.02	0.11	Credit ratings × Recession	−0.07	0.60
R^2	0.24		R^2	0.63	

E. Exploring Ultrahigh Dimensionality

In the preceding sections, we approach the reconstruction of numerical economic time series with linear models of topic attention time series. The view we propose, that the state of the economy is describable as time-varying configurations of topic attention, is likely to be better captured in regressions that allow for (potentially complex) interactions among topics. We explore this by extending the five-regressor lasso regressions for industrial production and employment in Table I to include not only the 180-topic attention time series as regressors, but also all pairwise topic interactions.

Table V reports the results. Interestingly, we find that no individual topics are selected for the model. Instead, *all* of the selected regressors are topic interactions, and the majority (8 of 10) include the recession topic. These results suggest that macroeconomic outcomes can be accurately captured not just by discussion of economic downturns, but also by specific attributes of downturns such as conditions in commodities markets and attention to corporate bankruptcy. More broadly, these results raise the possibility that more flexible nonlinear models from the machine learning toolkit, such as tree models that tend to be more adept than regressions for modeling covariate interaction, are likely to be valuable tools for measuring the state of the economy.

IV. News Attention and Macroeconomic Dynamics

Thus far we have analyzed contemporaneous correlations between news narratives of economic time series. This helps validate the conjecture that news can be a useful quantitative tool for summarizing the state of the economy and helps demonstrate the interpretability of quantitative analyses that use topic model estimates. However, it has little to say about whether news text conveys novel information that is distinct from information in standard numerical macroeconomic indicators, or whether news text helps in modeling longer-term macroeconomic trajectories.

A. Macroeconomic VAR

We investigate this question by studying the role of news attention in a macroeconomic VAR. We build on the five-variable monthly VAR specification studied by Baker, Bloom, and Davis (2016, BBD), which includes (in order) their EPU index, log value of the S&P 500 index, Federal Reserve funds rate, log employment, and log industrial production. The BBD VAR is a natural starting point for our analysis given the paper's large impact in macroeconomics and given that news text is an input to the EPU index.

In our baseline formulation, we augment the VAR to include news attention to the "recession" topic, based on its central role in the lasso regressions of Table I, in place of EPU (and we include three lags of all variables).[17] We estimate the VAR using data from 1985 to 2017.[18] We then plot output and employment impulse-response functions for a shock to "recession" news attention.[19] We define a news impulse as a shift in "recession" attention from its 5$^{\text{th}}$ to its 95$^{\text{th}}$ percentile. We use the same definition of an impulse for EPU in our comparisons below (following the impulse definition in BBD).

Panel A of Figure 4 plots the results. A shock to "recession" news attention generates a 1.99% drop in industrial production after 17 months. Employment declines by 0.92% after 20 months. These effects are highly statistically significant, as shown by the 90% confidence bands.[20] The estimated responses are large in economic magnitude. As a benchmark, Panel B compares our baseline VAR specification (solid black line) to the impulse response for EPU (solid gray line) in the VAR model of BBD.[21] The response of output to EPU (−0.98%) is roughly half as large as the response to "recession" attention, and the employment response (−0.34%) is roughly one-third as large.

Panel B demonstrates the robustness of our baseline VAR conclusions to alternative specifications. First, we consider ordering news attention last in the VAR while preserving the order of other covariates. In this case, the impulse responses reflect the macroeconomic impact of "recession" news coverage after controlling for contemporaneous responses to all other variables. This has a notable impact on our estimates, reducing the maximum output response from −1.99% to −1.25% (shown in gold, surrounded by the 90% confidence band also in gold). The maximum impact on employment is −0.51%, versus

[17] Section IV.E shows that this is a statistically optimal choice from the full set of topics.

[18] We update BBD's publicly available data (which run from 1985 to 2014) through June 2017 to correspond with our *WSJ* sample. We also replace their S&P 500 index data, which are constructed from the average daily index value within the month, with the value of the S&P 500 on the last day of the month. Using average daily values within the month causes additional serial correlation in log changes in the series (an increase from 6% with month-end values to 26% with average daily values). This change has only minor quantitative consequences for our results as well as the results of BBD.

[19] We define orthogonal shocks based on the Choleski decomposition given the BBD variable ordering plus the "recession" topic (ordered first in the vector).

[20] We compute bootstrap confidence intervals following Kilian (1999).

[21] This is the same five-variable specification of BBD, with EPU ordered first, and excluding "recession" attention.

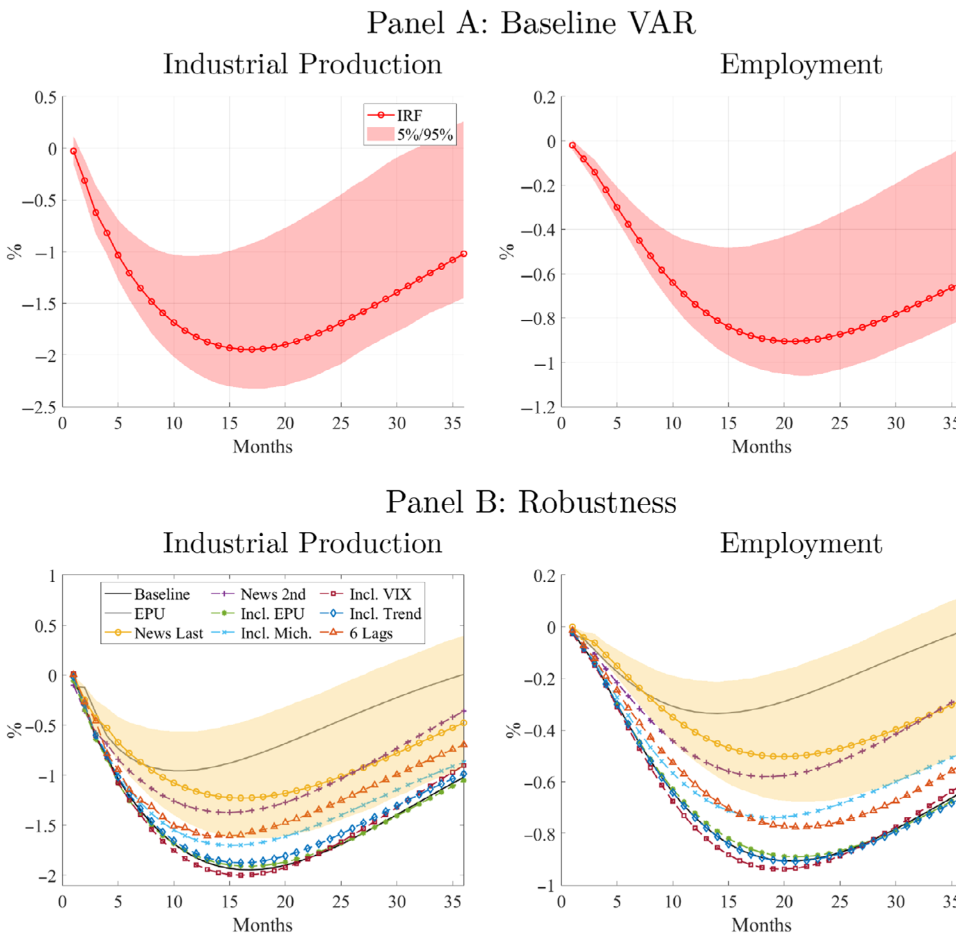


Figure 4. Output and employment responses to "recession" news attention shock. Responses of output and employment to a "recession" attention impulse in various VAR specifications, 1985 to 2017. The red shaded regions in the top plots are 90% confidence bands. This gold shaded regions in the bottom plots are 90% confidence bands when news attention is ordered last in the VAR (corresponding to the gold impulse-response curves). (Color figure can be viewed at wileyonlinelibrary.com)

−0.92% in the baseline specification. The effects of "recession" news attention, although smaller when it is ordered last in the VAR, remain highly significant and larger than the impulse responses to EPU.

If we instead alter the order of variables so that "recession" attention is second (after the S&P 500 index but before the Fed funds rate, employment, and industrial production), we see similar impulse responses as in the case in which news is ordered last. The difference in "baseline" and "news 2nd" specifications highlights that media outlets and financial markets have a partially synchronized response to news arrival (consistent with the high R^2 for the stock market return regression in Table I).

When we simultaneously control for EPU (which we order second in the VAR after "recession" attention), the "recession" attention impulse responses are essentially unchanged from the baseline. Replacing EPU with the Michigan Consumer Sentiment Index has a small effect on the output response to "recession" news (from −1.99% to −1.74%) and a small effect on the employment response (from −0.92% to −0.76%). Likewise, accounting for fluctuations in economic uncertainty via inclusion of the VIX index, introducing a time trend, and incorporating six rather than three lags in the VAR leave results qualitatively unchanged.

B. News Attention and Stock Market Dynamics

A large literature beginning with Fama (1990) argues that stock price fluctuations are driven by changing expectations about future macroeconomic outcomes. Section III.B documents a strong contemporaneous correlation between stock market returns and news attention (a correlation unmatched by any FRED-MD macro series). In light of this, a natural interpretation of this high correlation is that news text, like the stock market, reflects agents' expectations of the future macroeconomy, and these expectations are not well captured by time-series patterns in numerical macroeconomic data.

An intriguing implication of the VAR analysis is that recession attention and the stock market also contain notably *different* information about future macroeconomic outcomes. The top two panels of Figure 5 compare industrial production impulse responses for shocks to news attention and stock valuations.[22] The top left panel shows the baseline VAR (with news ordered first and the S&P 500 second) and the top right shows the VAR with S&P 500 ordered first and news second. Both cases tell a similar story: "recession" attention contains information about the future macroeconomy that is not fully reflected in stock prices. It has a large negative impact on future output, even when ordered after the S&P 500. Moreover, "recession" news has a *stronger* association with future output than the stock market does (regardless of ordering), which is a surprising and impressive fact given the Fama (1990) result.

The bottom-left panel of Figure 5 further illustrates the dynamic association between recession news and stock prices when news is ordered first. Upon arrival of a positive "recession" attention shock, prices initially drop by about 5%. But this response is incomplete. Over the next year prices drop further, reaching 7% below their preshock levels. When news is ordered second, the total response is smaller and borderline insignificant, but still economically large at −2.4% over 18 months.

[22] The S&P 500 impulse response is defined as the 95th percentile of log changes in the index. In other words, the shock corresponds to a one-month stock market return of 6.9%.

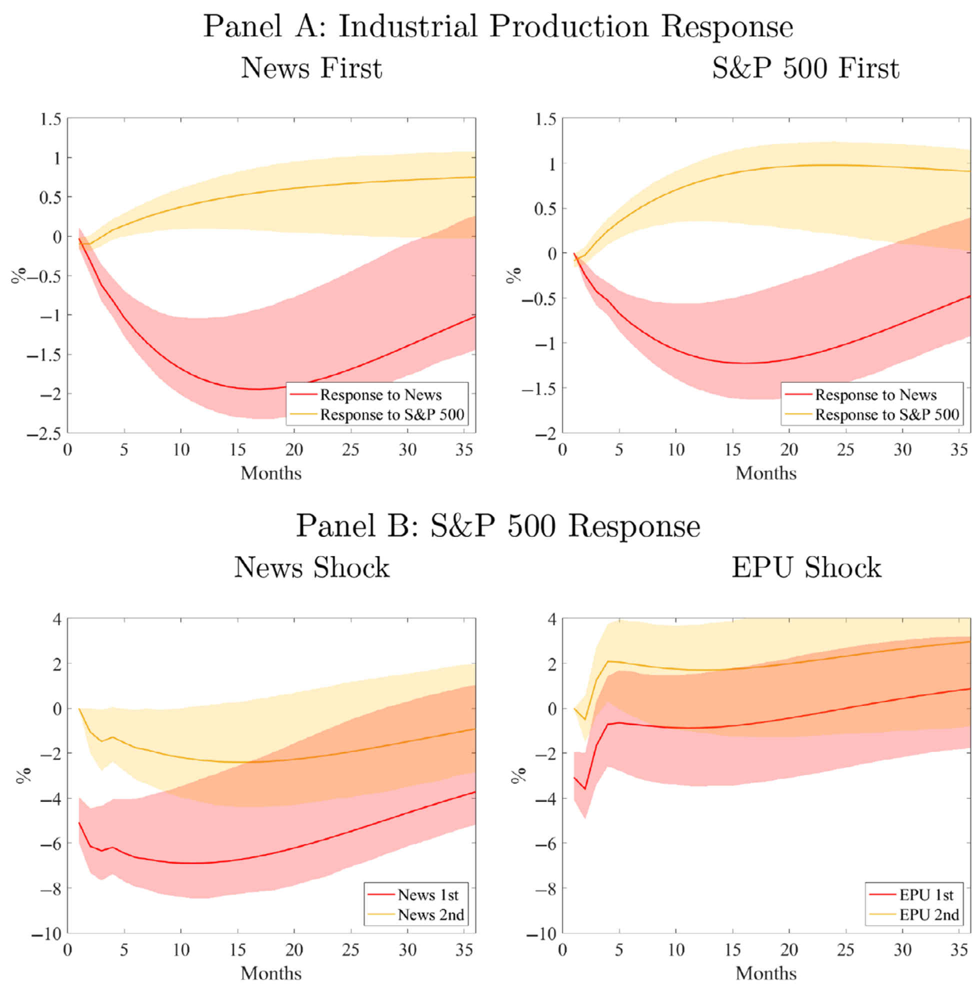


Figure 5. Stock market and "recession" attention. Panel A plots the response of output to impulses in "recession" attention and stock market value. Panel B reports the response of stock market value to an impulse in "recession" attention in our baseline VAR (left plot) and an impulse in EPU in the BBD VAR (right plot), and includes specifications for impulse variables ordered either before or after the S&P 500 in their respective VARs. Sample is 1985 to 2017. (Color figure can be viewed at wileyonlinelibrary.com)

The bottom right panel of Figure 5 helps illustrate the large economic magnitude of stock market responses to "recession" attention by comparing them to EPU impulses based on the VAR in BBD. When EPU is ordered first in the VAR, EPU has a negative and prolonged impact on stock prices, although the magnitude is about half as large as the impulse response to "recession" attention. The differences in these effects are statistically significant at the 90% level. Furthermore, if EPU is ordered second in the VAR (after the S&P 500), the stock market response to EPU is zero at first, and then becomes positive.

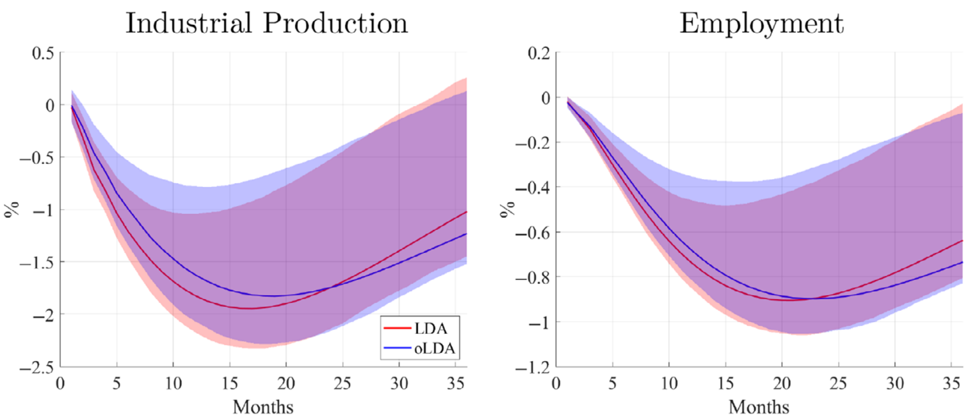


Figure 6. Responses to "recession" news attention shock: Avoiding look-ahead bias with online LDA. Responses of output and employment to a "recession" attention impulse for LDA versus oLDA, 1985 to 2017. The red lines represent the LDA point estimates and the red shaded regions 90% confidence bands. The blue lines represent the oLDA point estimates and the blue shaded regions 90% confidence bands. (Color figure can be viewed at wileyonlinelibrary.com)

C. Avoiding Look-Ahead Bias with oLDA

Next, we repeat our VAR analysis using the oLDA topic model estimates described in Section I.D. We extract the topic most similar to the "recession" topic in our main specification by comparing the top word lists and rerun the VAR exercise with the online version of the "recession" topic.

Figure 6 reports oLDA (no look-ahead) impulse-response estimates and compares them with the full-sample LDA impulse responses. We find that the effects of the oLDA "recession" news attention shock on both industrial production and employment are only slightly attenuated relative to the full-sample LDA estimates. For example, using the oLDA topic, output is 1.27% lower at 15 months, compared with 1.94% using full-sample LDA. Employment responses at 15 months decline from 0.86% with LDA to 0.76% with oLDA. This attenuation could result from either a reduction in look-ahead bias, or from the relative noisiness of oLDA estimates based on shorter samples. In any case, the differences are quantitatively minor. From a parameter uncertainty perspective, at short horizons LDA and oLDA estimates are distinct, but at longer horizons they are statistically indistinguishable.

D. Over a Century of Business News: 1890 to 2017

Our analysis thus far starts in 1984, which is when our full text sample starts. But the *WSJ* has been in print since 1889. A natural question is whether the conclusions we draw from the 1984 to 2017 sample extend back to earlier periods.

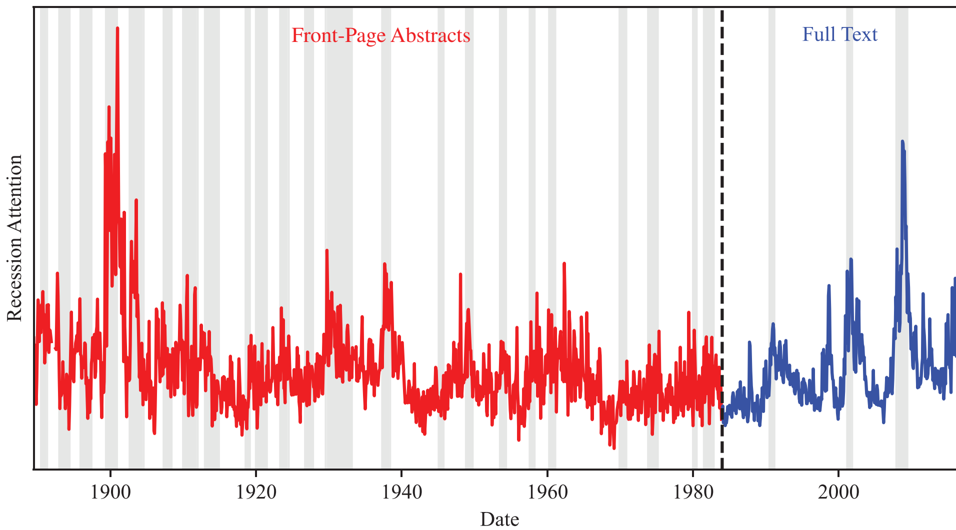


Figure 7. Recession attention and NBER recessions, 1890 to 2017. Time series of recession attention and NBER recessions, 1890 to 2017. The red line corresponds to the front-page abstracts sample and the blue line to the main full text sample. The gray shaded regions correspond to NBER recessions. (Color figure can be viewed at wileyonlinelibrary.com)

To answer this question, we leverage the sample of *WSJ* front-page titles and abstracts from Manela and Moreira (2017). Prior literature indicates that LDA does not perform well on short documents such as titles and abstracts (Tang et al. (2014)). Hence, we use the topic model learned from our main sample (full text articles from 1984 to 2017) to extend the topic attention series back to 1890.

To form estimates of attention using the abstracts corpus, we first constrain its vocabulary to that of the full text corpus. We then use the topic model trained on our main sample to calculate topic attention in the abstracts corpus from 1890 to 1983. Next, we append the fitted topics for the pre-1984 sample with the estimates from our main sample to arrive at a unified time series of topic attention from 1890 to 2017.

The major benefit of extending our sample is that we capture more than 20 recessions (vs. just three in our main sample). Figure 7 plots the time series of "recession" attention over this considerably longer sample. The unified series clearly shows that the *WSJ* allocates more space to recession coverage during NBER recessions. Interestingly, recession attention often rises before an NBER recession starts and declines before the recession ends. It also shows that recession attention estimates are noisier in the earlier sample of front-page abstracts compared to the later sample that includes full article text for all pages.

Given the longer attention series, we replicate our VAR analysis using data from 1939 to 2017. 1939 is the first year for which all of the VAR component

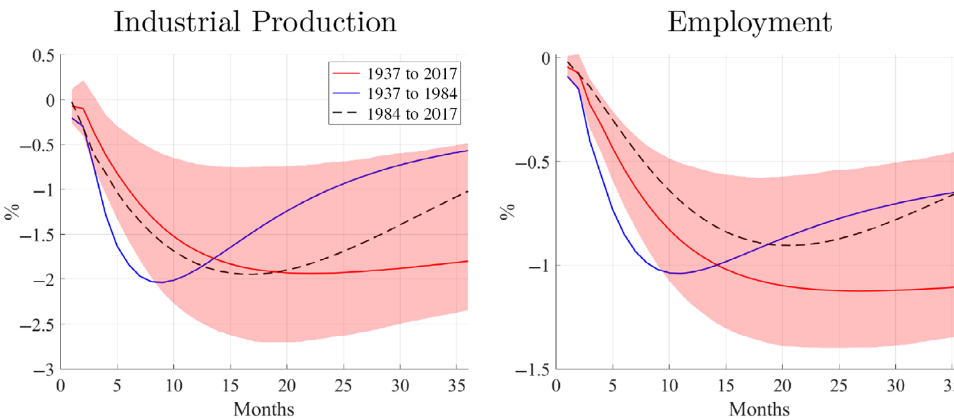


Figure 8. Responses to "recession" news attention shock: 1939 to 2017 sample. Responses of output and employment to a "recession" attention impulse, 1939 to 2017. We estimate news attention over the 1939 to 1984 period by applying our rich topic model to the *WSJ* front-page abstracts used by Manela and Moreira (2017). The shaded regions are 90% confidence bands. (Color figure can be viewed at wileyonlinelibrary.com)

series are available (nonfarm payroll employment has the shortest history, and begins in 1939). Figure 8 shows the impulse-response functions for these VAR estimates. Comparing the impulse responses to those in Figure 4, which are based on the more recent sample, two important patterns emerge. First, we find that the short-term response of both industrial production and employment to a recession news shock is quite similar across samples. In particular, 20 months after an increase in recession news attention, industrial production and employment decline by about 2% and 1%, respectively.

Second, we find that the long-run response estimates are more persistent in the longer sample. Estimates based on the more recent sample show considerable long-run reversion. Impulse responses of both output and employment revert to about half of their peak effects. By contrast, the longer front-page abstracts sample responses show little reversion at 36 months. For robustness, we also report VAR results based on the 1937 to 1984 sample and find similar estimates to those from our main 1984 to 2017 sample.

E. Attention Selection in a Text-Augmented VAR

Our VAR specification uses recession attention as the only news topic variable in the system. We next show that this is not an arbitrary choice, but is in fact a statistically optimal specification in terms of its trade-off of model fit and parameter parsimony.

Due to the large size of the topic model, including all news attention series in the VAR is obviously problematic due to "degrees-of-freedom problems" (Bernanke, Boivin, and Eliasz (2005)). Instead, we seek a VAR that includes news topics that genuinely influence macroeconomic dynamics and avoids spurious inclusion of irrelevant topics. This is a model selection problem, which we solve with a standard tool: cross-validated lasso regression.

We begin with the four core macroeconomic variables in the VAR specification above (S&P 500 index, Fed funds rate, employment, and industrial production), which we denote by y_t and fix as left-side variables in a multivariate regression. We then consider a collection of predictors that includes the 180 news attention series as well as EPU, VIX, and the Michigan consumer sentiment index, which we arrange as $x_t = (x_{1,t}, \ldots, x_{183,t})'$. The right-side variables are three lags each of y_t and x_t:

$$y_t = c + \sum_{l=1}^{3} (\Omega_l y_{t-l} + \Gamma_l x_{t-l}) + \epsilon_t. \tag{2}$$

Our modeling approach uses a lasso penalty to select robust predictors of y_t. We prefer to select or remove a right-side variable in its entirety (rather than selecting, say, only the third lag of a given variable, or selecting a variable to explain one element of y_t but not another). The group-lasso of Yuan and Lin (2006), recently used in finance by Freyberger, Neuhierl, and Weber (2020), is ideally suited for this task. Group-lasso assigns nonzero coefficients to only those predictor groups that most reliably forecast y_t, where the notion of "reliable" is determined by the penalty parameter λ. We use 183 groups, one for each variable in x_t (e.g., the group for $x_{j,t}$ includes all terms on the right-hand side of (2) associated with $x_{j,t-1}, x_{j,t-2}$, and $x_{j,t-3}$). We include lags of y_t without penalization.[23]

Figure 9 summarizes the importance of each predictor for forecasting macroeconomic outcomes. It shows the ℓ^2 norm of coefficients for each predictor variable as a function of the penalty parameter. When $\lambda \to 0$, there is no selection and all variables are included.[24] As λ rises, parameters become more heavily penalized and weak predictors begin to drop out. At high levels of λ, only the most potent macroeconomic predictors receive nonzero coefficients.

The legend reports the 10 predictor variables that survive heavy penalization, ordered in terms of statistical importance from top to bottom. One predictor stands out above all others—attention to "recession" news. Remarkably, its predictive contribution dominates EPU, VIX, and consumer sentiment (of these, only EPU makes the list of top 10 predictors). At lower levels of

[23] Specifically, our group-lasso estimation objective is

$$\min_{\{\Omega_l, \Gamma_l\}_{l=1}^{3}} \sum_{i=1}^{4} \sum_{t=5}^{T} \left(y_{i,t} - \sum_{j=1}^{4} \sum_{l=1}^{3} \omega_{i,j,l} y_{j,t-l} - \sum_{m=1}^{183} \sum_{l=1}^{3} \gamma_{i,m,l} x_{m,t-l} \right)^2 + \lambda \sum_{m=1}^{183} \sqrt{\sum_{j=1}^{4} \sum_{l=1}^{3} \gamma_{i,j,l}^2},$$

where $\omega_{i,j,l}$ and $\gamma_{i,j,l}$ are the (i, j) elements of Ω_l and Γ_l. All variables are variance-standardized prior to model estimation to ensure comparability of coefficient estimates.

[24] When $\lambda = 0$, the model has more regressors than observations and the estimator is not defined, and thus the first estimates on the left of Figure 9 correspond to a small positive value of λ.

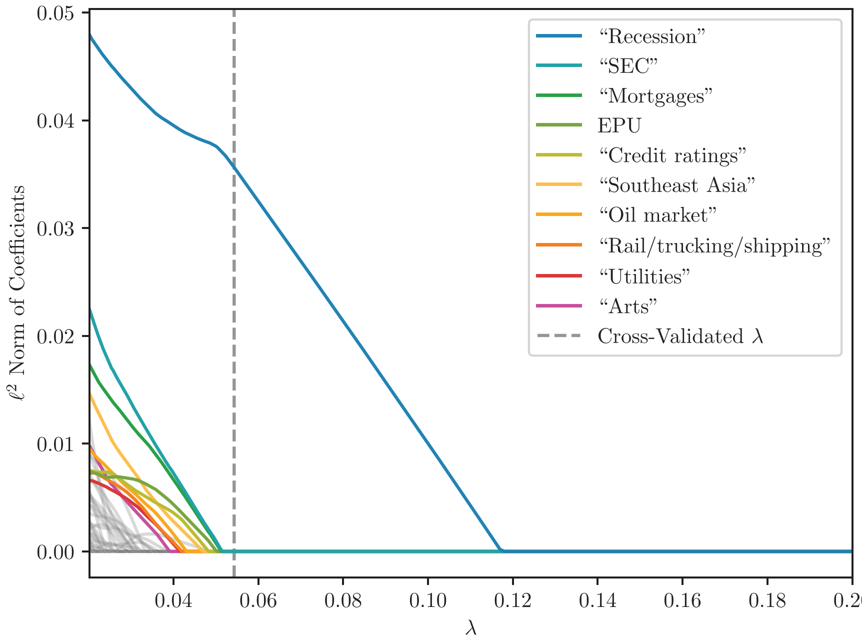


Figure 9. Group-lasso VAR selection. The ℓ^2 norm for all variables considered in a group-lasso VAR as a function of the penalty parameter λ. The legend shows the 10 predictors that survive the highest levels of penalization. Variable names in quotations refer to news attention series. (Color figure can be viewed at wileyonlinelibrary.com)

penalization, other news variables that enter the model include "SEC (Securities Exchange Commission)," "mortgages," and "credit ratings."

Figure 9 establishes that the recession topic is the appropriate choice for inclusion in the VAR if we were to choose just one topic. However, if we were to use the group-lasso to decide on a model specification, what is the best model? Equivalently, which value of λ is most appropriate? A standard criterion for selecting among λ values is cross-validation, which is designed to identify the model with the lowest expected out-of-sample forecast error. Specifically, we choose the λ corresponding to the model with lowest mean squared forecast error in 10-fold cross-validation.

The optimal cross-validated model corresponds to $\lambda = 0.054$, depicted as a vertical line in Figure 9. This model selects a single news attention topic, "recession," for inclusion in the VAR. It is based on this result that our text-augmented VAR specification uses "recession" attention alone, as it represents a statistically optimal set of news topics for VAR prediction.[25]

[25] A technical point regarding our construction of bootstrap standard errors in Figure 4 is that they are not adjusted for lasso model selection. In Internet Appendix Section IA.G, we describe and report impulse-response standard errors accounting for postselection inference. While postselection inference widens our confidence bands, the effect is small and does not change our

F. VAR Discussion

Combining macroeconomic analysis with topic modeling, we draw on a vast corpus of news text to better understand quantitative economic phenomena. Our VAR demonstrates a large incremental role for news attention in understanding economic dynamics. Our interpretation of this result is that news media reflect agents' perceptions of the state of the economy, including (but not limited to) their expectations of future macroeconomic conditions. It should not be surprising that news media can offer a representation of economic conditions that is unspanned by commonly studied numerical macroeconomic data. News contains information about our high-dimensional perceptions of the world around us, distilled into narratives through the complex process of human understanding.

Figure 4 provides evidence of economic fluctuations emerging from news. This news is an amalgamation of at least four phenomena. First, business news provides a summary of expectations about future productivity. The news-driven business cycle literature has made theoretical strides in understanding how changes in expectations induce booms and busts even absent concurrent changes in fundamentals (Beaudry and Portier (2004, 2007, 2014)).

Second is the closely connected role of noise in business cycle dynamics (as in Lorenzoni (2009), Angeletos and La'o (2010)). Inevitable imperfections in macroeconomic expectations (as summarized by news media) can induce volatility in economic activity. A third and also closely related component of news text is the sentiment of economic agents, akin to Keynesian "animal spirits," which drive fluctuations as in Angeletos and La'o (2013). Frequent *WSJ* interviews of influential asset managers, corporate executives, and policymakers are a likely vehicle through which sentiments become reflected in news text. Fourth, news production is determined by media firm incentives, which are influenced in turn by consumer demand (Mullainathan and Shleifer (2005), Gentzkow and Shapiro (2010)). As such, news text also reflects media slant, which can influence economic dynamics.

While we conduct our macroeconomic analysis in the context of a VAR, this is just one example of how to combine an economic model with news attention to aid inference and interpretation. Other exciting applications include using text narratives to better understand survey revisions and forecast errors (Coibion and Gorodnichenko (2015), Bordalo et al. (2020)) or to interpret residuals implied from DSGE models.

G. Narrative Retrieval

Empirical macroeconomics research has long wrestled with the issue of shock identification (see Ramey (2016) for a survey discussion). A common approach to model interpretation imposes identifying restrictions on the

conclusions about the statistical significance of news attention impulse responses. Hence, to simplify exposition in the main text, we stick with standard bootstrap confidence intervals without selection adjustment.

residual covariance matrix in a VAR, then studies impulse-response functions for each variable in the system. Our analysis in Figure 4 is one such example.

A great advantage of embedding news text in macroeconomic models is that fluctuations can be directly mapped to narratives of economic conditions in the underlying article text. This offers a new approach to model interpretation that complements traditional methods. Specifically, we propose a scheme for *narrative retrieval* in text-augmented VAR models. It relies on only two inputs—the estimated VAR coefficients and topic model parameters—without requiring the researcher to impose additional identifying restrictions or economic constraints.

First, we trace future macroeconomic outcomes to prevailing levels of news attention through the estimated VAR coefficients. From here, we further trace economic fluctuations to specific textual narratives using topic model estimates. In each month t, we identify the *WSJ* article published during the month that is most representative of a given VAR topic, k. This representative article is defined as the article i with the largest proportion of its content (θ_i) allocated to topic k.[26] This chain of estimates identifies the most influential individual articles—that is, specific narratives—that underlie model-implied macroeconomic expectations.

Our narrative retrieval procedure is a natural way to flag articles that coincide with large shifts in model predictions. It points the researcher to the most diagnostic articles for deep reading, bypassing the labor-intensive process of reading every document in the underlying corpus. Therefore, it automates the integration of narratives into macroeconomic models, rather than relying on manual narrative selection (as in Romer and Romer (1989)).

We demonstrate this approach in the baseline VAR of Figure 4, and study the narrative determinants of output expectations. In Figure 10, the red curve shows one-month-ahead expected output growth based on the estimated VAR. The black curve shows how "recession" news attention contributes to this expectation. This is another visualization of the large impact that news has on output dynamics to complement the impulse responses shown earlier.

Next, we annotate seven of the largest monthly drops in expected output growth coming from shifts in "recession" attention. Annotations show the headline of the article with the greatest attention allocation to "recession" at the time of the drop. These articles are the largest contributors to month-to-month variation in "recession" attention.

Reading the flagged articles carefully, we find that output growth expectations are underpinned by narratives that support theories discussed in Section IV.F. An informative example is the April 2001 article "Consumer Confidence Slides on Fears of Layoffs." The author, Greg Ip, notes that consumer confidence surveys have recently soured and offers an interpretation:

[26] Our VAR uses a single topic attention series, but this narrative retrieval approach generalizes to a VAR with any number of topic elements.

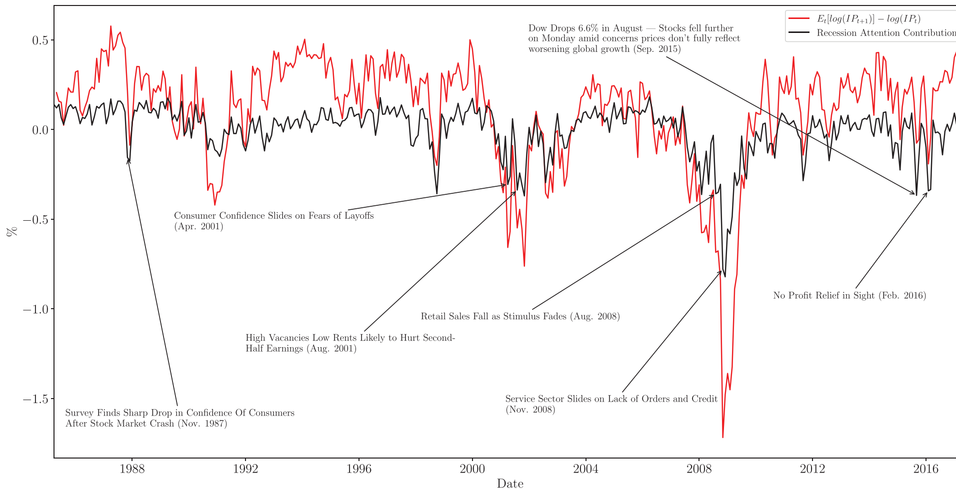


Figure 10. Narrative retrieval for industrial production growth forecasts. The figure shows one-month-ahead growth rate expectations from the baseline VAR (red line) and contribution to the expectation coming from "recession" attention (black line). We annotate seven of the largest shifts in "recession" attention-based expectations. Annotations are headlines of articles that have the greatest allocation to the "recession" topic in a given month. (Color figure can be viewed at wileyonlinelibrary.com)

> "mostly because consumers have been more pessimistic about the future as layoff announcements have mounted, energy costs have risen, and stock prices have fallen."

This is a narrative of deteriorating consumer demand arising from declining expectations for future growth, consistent with news-driven business cycle theory. The article then goes on to quote differing viewpoints of two experts, suggestive of the mechanism described by Angeletos and La'o (2013) in which macroeconomic expectations emerge from heterogeneous beliefs that propagate across agents.

A number of other articles also highlight narratives about economic growth expectations (November 2008, September 2015) and corporate profit expectations (August 2001, August 2008, February 2016) as primary drivers of VAR forecasts. The September 2015 article "Dow Drops 6.6% in August — Stocks Fell Further on Monday Amid Concerns Prices Don't Fully Reflect Worsening Global Growth" is another informative example. Why does news text provide such large incremental predictive information above S&P 500 valuations? This article suggests a rationale in the form of heterogeneous belief propagation. Despite the fact that "the Dow lost 6.6% in August, its largest one-month percentage decline since May 2010," the Dow fell another 0.7% on Monday, September 1, 2015. The narrative put forth by journalists Saumya Vaishampayan and Christopher Whittall is as follows:

> "Monday's selling was driven by intensifying concerns that stock prices don't fully reflect the deterioration in recent months of the global economic outlook, traders said.... While many traders and analysts remain upbeat on the prospects for the U.S. economy and generally optimistic about investing in the shares of large U.S. companies, market sentiment is broadly cautious following the sharp price swings of the past two weeks.... Some portfolio managers are heeding signs that growth is slowing in the economies of many developing nations and that economic gains are likely to be slow in coming elsewhere. 'Emerging markets don't look that great, the Federal Reserve is going to raise rates, and Europe is still gradually emerging from recession,' said David Lebovitz, global market strategist for J.P. Morgan Asset Management."

The November 1987 article "Survey Finds Sharp Drop in Confidence of Consumers after Stock Market Crash" is yet another informative data point. This article's narrative illustrates how the sudden and extreme stock price drop in October 1987 dented consumer confidence:

> "Consumer Sentiment stood at 92.5 before the panic selloff on Wall Street Oct. 19 but plummeted to 82.4 for the rest of the month.... 'Needless to say the current situation is volatile,' the surveyors said. 'Whether a recession eventually develops can't be determined from these data (because) the response by consumers has yet to move past their initial reactions. Nonetheless the data indicate weakening sales and a much larger potential downside risk.'... The surveyors conceded that 'caution must be used in the interpretation of these results because the post-crash sample was so small.'... Mr. Curtin[27] said consumers still haven't assessed the personal toll such as reduced income or lost jobs that the crash may ultimately exact."

In the context of the text-augmented VAR, this narrative is associated with a drop in expected one-month industrial production growth of roughly 0.25%. The causes of the 1987 crash are hotly debated to this day,[28] but it is generally agreed that the crash was a transitory incident with little effect on subsequent macroeconomic fundamentals (Barro (1990)). All the more impressive, then, is how the uncertainty emphasized in this article reduces economic expectations in the immediate aftermath of the crash. The decline in VAR-based output expectations recovers within two months of the crash as recession narratives associated with the crash quickly disappear. This episode illustrates the effects of a prototypical noise shock in noisy business theories (Lorenzoni (2009), Angeletos and La'o (2010)).

[27] Richard T. Curtin, director of consumer surveys at the University of Michigan's Institute of Social Research.

[28] Potential explanations range from the prevalence of portfolio insurance strategies (Brady (1988)) to proposed and eventually abandoned legislation on corporate takeovers (Mitchell and Netter (1989)).

V. Stock Market Timing

In this section, we investigate the association between news attention and aggregate market return dynamics. Predictable variation in returns may arise from at least two sources, namely, market inefficiency and time-varying discount rates. In inefficient capital markets, some information contained in the newspaper may not be fully priced by the stock market (Fama (1970)). Predictability can also arise if news attention to specific topics is indicative of systematic behavioral biases of investors (Shiller (2003)). Alternatively, even in efficient markets, rational time variation in discount rates can lead to return predictability (Cochrane (2011)). For example, if heightened attention to recession news indicates an elevated disaster probability, such periods should be followed by higher expected returns (e.g., Wachter (2013)).

When considering return predictability based on news attention, we are cognizant that full-sample topic estimates are susceptible to look-ahead bias. To address this, in addition to our full-sample LDA attention estimates, we also consider a time series of oLDA topic attention estimates (introduced in Section IV.C) that is free of look-ahead bias. When using oLDA we use the same number of topics as in the full-sample LDA specification.

For both topic models, we form a lasso-based forecast of market returns selecting five regressors as in the main specifications. We detrend monthly attention estimates because, as Figure 3 shows, many topics such as "recession" and "terrorism" are increasingly covered by the newspaper. Following Kelly, Malamud, and Zhou (2024), returns are standardized by their trailing 12-month standard deviation, while we standardize the signals by the trailing full-sample standard deviation.

We train the lasso model out-of-sample starting with half the sample as training data and then evaluate one-month out-of-sample before extending the training sample with the new month of data and repeating the procedure. This expanding sample procedure results in 174 months of test sample data. We start our sample in 1988—excluding the first four years from the sample as oLDA requires a burn-in period to converge. This period corresponds to the point at which the log-likelihood estimates for oLDA stabilize.

To evaluate the economic significance of the forecasting model, we convert our return forecasts into an out-of-sample market-timing strategy. Each month t, the strategy takes a position in the market that is proportional to the one-month-ahead out-of-sample return forecast as of month t.

We compare our news-based timing strategies to several natural benchmarks. One is a naive buy-and-hold strategy that holds the market in constant proportion throughout our sample. We also consider trading strategies for two alternative text-based measures: (i) the frequency of mentions of the word "recession" in each month, inspired by *The Economist*'s R-word index (ii) the economic policy index used previously in Section IV.A. We detrend both of these text-based measures in the same way we detrend the LDA attention series. Finally, we consider a trading strategy using the 15 monthly predictors from the "kitchen sink" forecasting regression in GW. For all but

Table VI
Out-of-Sample Performance of Market-Timing Strategies

This table reports annualized performance measures for the different market-timing strategies. Buy-and-hold corresponds to a strategy that holds the market. LDA corresponds to a lasso strategy using our original full-sample LDA attention series. oLDA corresponds to a lasso strategy that uses oLDA in-place of full-sample LDA. R-word uses a prediction based on mentions of the word "recession" in the *WSJ*. EPU uses a prediction based on the EPU index. GW uses the 15 predictors from the Welch and Goyal (2008) "kitchen sink" regression. GW Lasso uses a similar lasso procedure as oLDA with the same 15 predictors from Welch and Goyal (2008). All strategies except buy-and-hold, hold the forecasted market return to the following month.

	Buy-and-Hold	LDA	oLDA	R-Word	EPU	GW	GW Lasso
Sharpe ratio	0.71	0.99	1.04	0.65	0.53	0.53	0.62
Expected return	2.72	3.79	4.01	2.50	2.04	2.03	2.40
t-Stat	2.69	3.76	3.97	2.48	2.02	2.01	2.37
Information ratio		0.29	0.31	−0.46	−0.50	−0.18	−0.07
α		2.30	2.87	−0.20	−0.51	0.61	1.55
α t-Stat		2.67	3.06	−1.54	−1.43	0.69	1.58
Max loss	−47.09	−47.31	−25.89	−46.98	−77.66	−65.76	−64.31
Skew	−0.56	0.20	1.04	−0.66	−1.41	−0.71	0.05

the buy-and-hold strategy, we regress next month's stock market returns on the predictors during a training period. We then form a predicted return $\hat{r}_{t+1}$ based on month t predictors. Finally, we hold $\hat{r}_{t+1}$ units of the index to the next period, yielding a trading strategy return of $\hat{r}_{t+1} \times r_{t+1}$. In the case of the GW kitchen sink regression, we also consider a lasso-based timing strategy in analogy to the lasso regression step underlying the news attention trading strategy.

A. Results

Table VI reports the main results on market timing with news attention. We report annualized performance statistics for the different strategies, which are commonly used to compare strategies by the market-timing literature (e.g., Kelly, Malamud, and Zhou (2024)). For each strategy, we report the Sharpe ratio, expected return, and corresponding t-statistic for the expected return.

To study the excess performance of each strategy relative to the buy-and-hold benchmark, we next report the information ratio (IR), alpha, and alpha t-statistic from a time-series regression of each strategy on the buy-and-hold returns. Finally, we also report the maximum loss, or the lowest return earned by the strategy, and its skewness.

We find that news attention provides a useful signal for market timing. Both the LDA and oLDA strategies generate the highest Sharpe ratio, alpha, and t(alpha). At the same time, its maximum loss is modestly elevated but comparable to the other benchmarks.

The worst performer is the GW kitchen sink regression model. Because the advantage of LDA over the GW strategy could be due to its use of

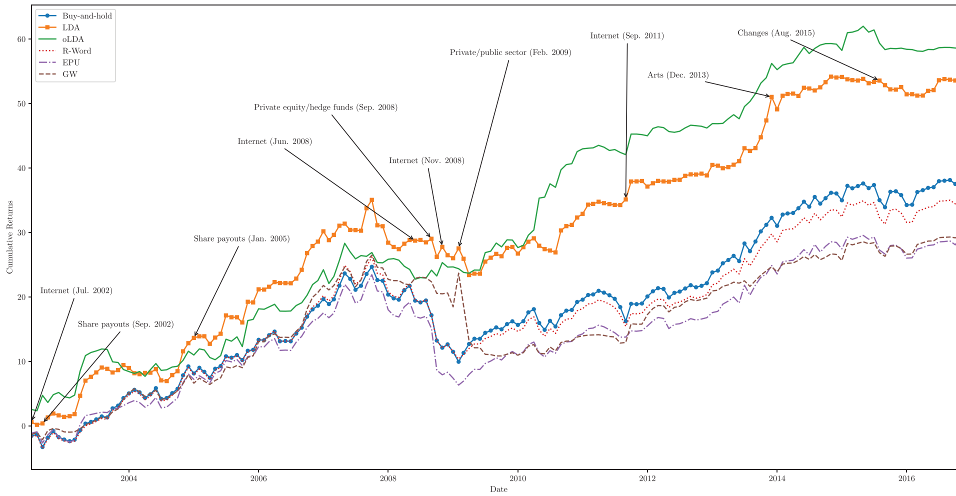


Figure 11. Cumulative returns for market-timing strategies. This figure plots the cumulative returns for different market-timing strategies over the testing sample. Buy-and-hold corresponds to a strategy that holds the market with static leverage. LDA corresponds to a lasso strategy using our original full-sample LDA attention series. oLDA corresponds to a lasso strategy that uses online LDA in-place of full-sample LDA. R-word uses a prediction based on mentions of the word "recession" in the *WSJ*. EPU uses a prediction based on the EPU index. GW uses the 15 predictors from the Welch and Goyal (2008) "kitchen sink" regression. Returns are standardized with the same standard deviation as the buy-and-hold strategy for interpretability. Labeled are the months with the largest LDA outperformance relative to the buy-and-hold benchmark and their most relevant topic. (Color figure can be viewed at wileyonlinelibrary.com)

regularization, we also analyze a lasso variant similar to the one we use with LDA. As the table shows, performance of the regularized GW strategy remains below that of the buy-and-hold benchmark, and well below that of LDA.

The ability of our LDA strategy to outperform the market suggests that paying attention to *WSJ* coverage can be quite lucrative for investors. It also shows that, despite reducing the dimensionality of the newspaper text in an unsupervised manner, the reduction effectively summarizes the main themes of the newspaper. This reduction enables the forecasting regression model to estimate expected market returns based on high-level topics, rather than specific word usage such as "recession."

B. Narrative Retrieval

Figure 11 shows the cumulative returns for the different strategies. It reveals the points in time when LDA diverges by outperforming the benchmark strategies. For example, while holding the market or pursuing most other strategies would have generated substantial losses in 2008 and 2011, the LDA strategy reduces its market exposure during these periods.

What does the *WSJ* report during these periods that leads the LDA strategy to expect below-average market returns? We next run a variant of the above narrative retrieval analysis but for the market-timing application. We aim to understand the news shocks that correspond most to large differences between our market-timing strategy's performance and that of the simple buy-and-hold strategy.

To retrieve the narratives underlying these performance gains, we label the months in Figure 11 with the largest performance improvement in our timing strategy over the buy-and-hold strategy. We denote the change in topic k's attention at time t by $\Delta\theta_{t,k}$, the regression coefficient estimated over the preceding training period by $\hat{\beta}_{t,k}$, and the realized market return by r_t. We select the topic k, which contributes most to the market-timing strategy's performance in month $t+1$, as follows:

$$\hat{k}_{t+1} = \text{argmax}_k \Delta\theta_{t,k}\hat{\beta}_{t,k}r_{t+1}. \tag{3}$$

We can then isolate the articles most relevant to this improvement by sorting articles based on their exposure to this topic.

For example, our predictive model identifies increased attention to the "private/public sector" topic as a strong negative signal about future market returns because prior such news preceded stock market declines. This leads the LDA strategy to short the market going into February 2009, and to generate a three-standard-deviation return gap relative to the declining market, which is the largest gap in our sample. By focusing on the articles with the largest attention to the private/public-sector topic, we can discern a clear narrative for this outperformance.

In January 2009, the newly elected Obama administration frantically tries to restore confidence in the U.S. banking sector. It also initiates its key health care reform, which would later become the Affordable Care Act. For example, the *WSJ* reports under the headline "Stimulus Package Unveiled — $825 Billion Plan Includes Business Tax Breaks; Senate Releases Cash for Bank Rescues" that

> "The stimulus plan was released hours before the Senate backed Mr. Obama on another key measure by approving the Treasury's call for the release of the second half of the 700 billion financial system rescue package. That program is being replenished as Bank of America Corp. was near an agreement with U.S. officials that would provide it with 15 billion to 20 billion of fresh capital. The plan would be one of the largest single government expenditures in U.S. history and would be equivalent to about 3 of gross domestic product over two years."

Later that month, under the headline "Banks Hit by Nationalization Fears — Financials Plunge As U.S. Considers New Rescue Options," the *Journal* reports that

> "Shares of the biggest names in American banking plunged Tuesday as some investors feared that the government would need to nationalize the most deeply wounded financial institutions wiping out stockholders. The hours-old administration of President Barack Obama is expected to move

> swiftly to try to stabilize the financial system by pumping more capital into weakened banks and buying bad assets. Nationalization appears to be a last resort but other options on the table move the U.S. in that direction."

In February 2009, when this outperformance realizes, confidence in banks worsens and many believe the U.S. government is destined to take over some of the largest banks. The *WSJ* reports under the headline "U.S. Seeks to Stem Bank Fears — White House Plays Down Nationalization Talk as Stocks Hit New Bear-Market Low" that

> "The White House tried to knock down speculation that the government is preparing to nationalize several large U.S. banks, but some bankers complained that the Obama administration needs to act even more aggressively to shore up confidence in battered financial institutions."

The budget proposal by the new administration generally promises a larger role for the federal government in the U.S. economy, which likely worsens such fears and further depresses market prices. A front-page article entitled "Obama Budget Pushes Sweeping Change" describes this policy change as follows:

> "The budget blueprint for fiscal year 2010 is one of the most ambitious policy prescriptions in decades, a reordering of the federal government to provide national health care, shift the energy economy away from oil and gas, and boost the federal commitment to education."

These articles suggest that the notable outperformance in February 2009 stems from an underestimation by investors of the degree to which the financial crisis would prompt the new administration to redraw the boundaries between the private and public sectors of the U.S. economy, and what this policy change would imply for shareholders.

VI. Conclusion

Understanding the forces that drive fluctuations in the state of the economy is central to economic modeling. The overwhelming majority of empirical research approached this problem by focusing on numerical macroeconomic indicators. We offer an alternative approach that summarizes economic conditions in terms of narratives in business news.

Our approach is motivated by the view that news text is a reflection of the state of the economy. The media sector is an information intermediary that meets the information demand of consumers and investors with verbal descriptions of economic events and their interpretation. Of course, news media is a verbal mirror of the economy that comes with flawed refractions, often in the form of producer and consumer biases, noisy inferences, and speculative sentiment.

We estimate a topic model from the full text of the *WSJ*. Our estimates provide a taxonomy of themes that are subjects of attention in financial markets and the broader economy. We measure the amount of attention allocated to each theme at each point in time, and then use these measurements as inputs into statistical models of economic fluctuations.

Economic topics identified from the *WSJ* text closely coincide with related numerical measures of economic activity, including macroeconomic aggregates like output and employment, financing activity, asset prices, and different measures of economic uncertainty. In the context of a standard macroeconomic VAR, we show that news topic attention strongly influences economic dynamics above and beyond standard numerical economic indicators. In the context of stock market timing, we show that an oLDA-based strategy outperforms the market as well as several natural alternative strategies.

We propose a new perspective on model interpretation with narrative retrieval. Our approach relies on the model to digest massive text corpora that are beyond human readability and flags articles that are most statistically related to a specific fluctuation of interest. It thus isolates news-based narratives of those events that the researcher can "close-read" for detail and nuance about the drivers of model-based expectations.

Finally, readers may explore our results in greater detail through the interactive website, www.structureofnews.com , where we post all of our topic model output. Through the website, researchers can download our *WSJ* news attention time series for use in their own projects. We intend to post regular model updates as new data become available.

Initial submission: April 14, 2022; Accepted: May 15, 2023
Editors: Stefan Nagel, Philip Bond, Amit Seru, and Wei Xiong

REFERENCES

Angeletos, George-Marios, and Jennifer La'o, 2010, Noisy business cycles, *NBER Macroeconomics Annual* 24, 319–378.

Angeletos, George-Marios, and Jennifer La'o, 2013, Sentiments, *Econometrica* 81, 739–779.

Baker, Scott R., Nicholas Bloom, and Steven J. Davis, 2016, Measuring economic policy uncertainty, *Quarterly Journal of Economics* 131, 1593–1636.

Baker, Scott R., Nicholas Bloom, Steven J. Davis, and Marco C. Sammon, 2021, What triggers stock market jumps? Technical report, National Bureau of Economic Research.

Barro, Robert J., 1990, The stock market and investment, *Review of Financial Studies* 3, 115–131.

Beaudry, Paul, and Franck Portier, 2004, An exploration into Pigou's theory of cycles, *Journal of Monetary Economics* 51, 1183–1216.

Beaudry, Paul, and Franck Portier, 2007, When can changes in expectations cause business cycle fluctuations in neo-classical settings? *Journal of Economic Theory* 135, 458–477.

Beaudry, Paul, and Franck Portier, 2014, News-driven business cycles: Insights and challenges, *Journal of Economic Literature* 52, 993–1074.

Bernanke, Ben S., Jean Boivin, and Piotr Eliasz, 2005, Measuring the effects of monetary policy: A factor-augmented vector autoregressive (FAVAR) approach, *Quarterly Journal of Economics* 120, 387–422.

Blei, David M., Andrew Y. Ng, and Michael I. Jordan, 2003, Latent Dirichlet allocation, *Journal of Machine Learning Research* 3, 993–1022.

Bordalo, Pedro, Nicola Gennaioli, Yueran Ma, and Andrei Shleifer, 2020, Overreaction in macroeconomic expectations, *American Economic Review* 110, 2748–2782.

Boudoukh, Jacob, Ronen Feldman, Shimon Kogan, and Matthew Richardson, 2018, Information, trading, and volatility: Evidence from firm-specific news, *Review of Financial Studies* 32, 992–1033.

Boydstun, Amber E., 2013, *Making the News: Politics, the Media, and Agenda Setting* (University of Chicago Press, Chicago).

Brady, Nicholas, 1988, *Report of the Presidential Task Force on Market Mechanisms* (U.S. Government Printing Office, Washington, D.C).

Campbell, John Y., and Samuel B. Thompson, 2008, Predicting excess stock returns out of sample: Can anything beat the historical average? *Review of Financial Studies* 21, 1509–1531.

Chahrour, Ryan, Kristoffer Nimark, and Stefan Pitschner, 2021, Sectoral media focus and aggregate fluctuations, *American Economic Review* 111, 3872–3922.

Chauvet, Marcelle, and Jeremy Piger, 2008, A comparison of the real-time performance of business cycle dating methods, *Journal of Business & Economic Statistics* 26, 42–49.

Cochrane, John H., 2011, Presidential address: Discount rates, *Journal of Finance* 66, 1047–1108.

Coibion, Olivier, and Yuriy Gorodnichenko, 2015, Information rigidity and the expectations formation process: A simple framework and new facts, *American Economic Review* 105, 2644–2678.

Cong, Lin William, Tengyuan Liang, and Xiao Zhang, 2020, Textual factors: A scalable, interpretable, and data-driven approach to analyzing unstructured information, Working paper, Cornell University.

Ellingsen, Jon, Vegard Larsen, and Leif Anders Thorsrud, 2020, News media vs. FRED-MD for macroeconomic forecasting, Working Paper ID 3720396, Social Science Research Network, Rochester, NY.

Engle, Robert F., 2019, Financial volatility in an age of geopolitical risks, Technical report, NYU Working paper.

Engle, Robert F., Stefano Giglio, Bryan T. Kelly, Heebum Lee, and Johannes Stroebel, 2020, Hedging climate change news, *Review of Financial Studies* 33, 1184–1216.

Engle, Robert F., Takatoshi Ito, and Wen-Ling Lin, 1990, Meteor showers or heat waves? Heteroskedastic intra-daily volatility in the foreign exchange market, *Econometrica* 58, 525–542.

Fama, Eugene F., 1970, Efficient capital markets: A review of theory and empirical work, *Journal of Finance* 25, 383–417.

Fama, Eugene F., 1990, Stock returns, expected returns, and real activity, *Journal of Finance* 45, 1089–1108.

Freyberger, Joachim, Andreas Neuhierl, and Michael Weber, 2020, Dissecting characteristics nonparametrically, *Review of Financial Studies* 33, 2326–2377.

García, Diego, 2013, Sentiment during recessions, *Journal of Finance* 68, 1267–1300.

Gentzkow, Matthew, Bryan Kelly, and Matt Taddy, 2019, Text as data, *Journal of Economic Literature* 57, 535–574.

Gentzkow, Matthew, and Jesse M. Shapiro, 2010, What drives media slant? Evidence from U.S. daily newspapers, *Econometrica* 78, 35–71.

Hansen, Stephen, Michael McMahon, and Andrea Prat, 2017, Transparency and deliberation within the FOMC: A computational linguistics approach, *Quarterly Journal of Economics* 133, 801–870.

Herskovic, Bernard, Bryan Kelly, Hanno Lustig, and Stijn Van Nieuwerburgh, 2016, The common factor in idiosyncratic volatility: Quantitative asset pricing implications, *Journal of Financial Economics* 119, 249–283.

Hoffman, Matthew, Francis Bach, and David Blei, 2010, Online learning for latent Dirichlet allocation, *Advances in Neural Information Processing Systems* 23, 856–864.

Ke, Zheng Tracy, Bryan T. Kelly, and Dacheng Xiu, 2019, Predicting returns with text data, Technical report, National Bureau of Economic Research.

Kelly, Bryan T., Asaf Manela, and Alan Moreira, 2021, Text selection, *Journal of Business & Economic Statistics* 39, 1–21.

Kelly, Bryan T., Semyon Malamud, and Kangying Zhou, 2024, The virtue of complexity in return prediction, *Journal of Finance* 79, 459–503.

Kilian, Lutz, 1999, Finite-sample properties of percentile and percentile-t bootstrap confidence intervals for impulse responses, *Review of Economics and Statistics* 81, 652–660.

Larsen, Vegard H., and Leif A. Thorsrud, 2019, The value of news for economic developments, *Journal of Econometrics* 210, 203–218.

Lorenzoni, Guido, 2009, A theory of demand shocks, *American Economic Review* 99, 2050–2084.

Manela, Asaf, and Alan Moreira, 2017, News implied volatility and disaster concerns, *Journal of Financial Economics* 123, 137–162.

McCracken, Michael W., and Serena Ng, 2016, FRED-MD: A monthly database for macroeconomic research, *Journal of Business & Economic Statistics* 34, 574–589.

Mitchell, Mark L., and Jeffry M. Netter, 1989, Triggering the 1987 stock market crash: Antitakeover provisions in the proposed house ways and means tax bill? *Journal of Financial Economics* 24, 37–68.

Mullainathan, Sendhil, and Andrei Shleifer, 2005, The market for news, *American Economic Review* 95, 1031–1053.

Murtagh, Fionn, and Pierre Legendre, 2014, Ward's hierarchical agglomerative clustering method: Which algorithms implement Ward's criterion? *Journal of Classification* 31, 274–295.

Quinn, Kevin M., Burt L. Monroe, Michael Colaresi, Michael H. Crespin, and Dragomir R. Radev, 2010, How to analyze political attention with minimal assumptions and costs, *American Journal of Political Science* 54, 209–228.

Ramey, V.A., 2016, Macroeconomic shocks and their propagation, in *Handbook of Macroeconomics*, Vol. 2, 71–162 (Elsevier, Amsterdam).

Romer, Christina D., and David H. Romer, 1989, Does monetary policy matter? A new test in the spirit of Friedman and Schwartz, *NBER Macroeconomics Annual* 4, 121–170.

Shiller, Robert J., 2003, From efficient markets theory to behavioral finance, *Journal of Economic Perspectives* 17, 83–104.

Tang, Jian, Zhaoshi Meng, Xuanlong Nguyen, Qiaozhu Mei, and Ming Zhang, 2014, Understanding the limiting factors of topic modeling via posterior contraction analysis, International Conference on Machine Learning, 190–198.

Thorsrud, Leif Anders, 2020, Words are the new numbers: A newsy coincident index of the business cycle, *Journal of Business & Economic Statistics* 38, 393–409.

Tibshirani, Ryan J., Jonathan Taylor, Richard Lockhart, and Robert Tibshirani, 2016, Exact post-selection inference for sequential regression procedures, *Journal of the American Statistical Association* 111, 600–620.

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

Welch, Ivo, and Amit Goyal, 2008, A comprehensive look at the empirical performance of equity premium prediction, *Review of Financial Studies* 21, 1455–1508.

Xiao, Han, and Thomas Stibor, 2010, Efficient collapsed Gibbs sampling for latent Dirichlet allocation, Proceedings of 2nd Asian Conference on Machine Learning, 63–78, JMLR Workshop and Conference Proceedings.

Yuan, Ming, and Yi Lin, 2006, Model selection and estimation in regression with grouped variables, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 68, 49–67.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Appendix S1: Internet Appendix.
Replication Code.