FEDERAL RESERVE BANK OF SAN FRANCISCO

WORKING PAPER SERIES

Measuring News Sentiment

Adam Hale Shapiro
Federal Reserve Bank of San Francisco

Moritz Sudhof
Kanjoya

Daniel Wilson
Federal Reserve Bank of San Francisco

January 2017

Working Paper 2017-01
http://www.frbsf.org/economic-research/publications/working-papers/wp2017-01.pdf

Suggested citation:

Shapiro, Adam Hale, Moritz Sudhof, Daniel Wilson. 2017. "Measuring News Sentiment" Federal Reserve Bank of San Francisco Working Paper 2017-01.
http://www.frbsf.org/economic-research/publications/working-papers/wp2017-01.pdf

Measuring News Sentiment*

Adam Hale Shapiro[†], Moritz Sudhof[‡] and Daniel Wilson[§]

January 4, 2017

Abstract

We develop and assess new time series measures of economic sentiment based on computational text analysis of economic and financial newspaper articles from January 1980 to April 2015. The text analysis is based on predictive models estimated using machine learning techniques from Kanjoya. We analyze four alternative news sentiment indexes. We find that the news sentiment indexes correlate strongly with contemporaneous business cycle indicators. We also find that innovations to news sentiment predict future economic activity. Furthermore, in most cases, the news sentiment measures outperform the University of Michigan and Conference board measures in predicting the federal funds rate, consumption, employment, inflation, industrial production, and the S&P500. For some of these economic outcomes, there is evidence that the news sentiment measures have significant predictive power even after conditioning on these survey-based measures.

*We thank Armen Berjikly and the Kanjoya staff for generously assisting on the project and providing guidance, comments and suggestions. Kevin Vo and Catherine van der List provided excellent research assistance. The views expressed in this paper are solely those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of San Francisco, or the Board of Governors of the Federal Reserve System.

†Federal Reserve Bank of San Francisco, adam.shapiro@sf.frb.org

‡Kanjoya, moritz@cs.stanford.edu, moritz_sudhof@ultimatesoftware.com

§Federal Reserve Bank of San Francisco, daniel.wilson@sf.frb.org

1 Introduction

Policymakers and market participants care a great deal about current and future aggregate business conditions. The general nowcasting and forecasting toolkit relies on a broad array of models that incorporate both "hard" and "soft" information. The former includes objective and directly quantifiable variables such as production and employment, while the latter includes more subjective variables typically constructed from survey responses concerning attitudes about current and future economic conditions. There are a broad array of soft variables available, but the survey-based indexes of consumer sentiment by the University of Michigan and the Conference Board are the most widely followed. These measures have been shown to have important predictive power, helping to forecast macroeconomic outcomes even after controlling for a host of factors (Souleles (2004), Carroll, Fuhrer, and Wilcox (1994), Bram and Ludvigson (1998)).

Two primary mechanisms have been posited for how economic sentiment affects economic activity. First, there is the "animal spirits" hypothesis, whereby consumer and business sentiment itself can drive economic activity. For instance, a consumer sentiment shock has been blamed as the culprit for the 1990−91 recession (Blanchard (1993)). A number of recent papers argue that "sentiment shocks" can drive aggregate business conditions (Angeletos and La'O (2013), Benhabib, Wang, and Wen (2015)). Second, sentiment may be purely informational, containing news about future states of the economy held by the public but not (yet) observed in the hard data. Indeed, Barsky and Sims (2012) found that this informational component forms the main link between sentiment and future activity. While the debate as to *why* sentiment and future economic activity may be related is still ongoing, the consensus among researchers is that a correlation between sentiment and future economic activity does exist and is therefore useful for forecasting purposes.

In this study, we develop and assess an index designed to measure the economic sentiment embodied in the news media. Unlike survey-based measures of economic sentiment, our index relies on computational text analysis to extract sentiment from economic and financial newspaper articles. Text-based measures of economic activity are becoming more popular among researchers due to their apparent advantages over surveys in terms of cost and scope (Fraiberger (2016)). Surveys are inherently expensive to conduct, oftentimes based on a relatively small samples of individuals, and therefore may be subject to sampling problems

(Ludvigson (2004)).

The novelty of our method is to use a proprietary machine learning predictive model, developed by the company Kanjoya,[1] which scores text in terms of the degree to which it expresses each of a large number of different emotions. This model assesses entire sentence structures and assigns a probability score as to whether the sentence falls into a certain emotional category. The model is trained on data compiled from Kanjoya's social networking site, the Experience Project (EP). Users on the EP site discuss personal experiences and feelings (text) while also identifying (tagging) the emotions they are feeling and soliciting emotional reactions from other users. This provides a mapping from text to emotional labels that Kanjoya uses to train a predictive sentiment model via structured machine learning. Research samples from the EP corpus have been used previously in the Natural Language Processing literature.[2]

This machine learning approach is in contrast to the commonly used lexical or "bag-of-words" approach. In that approach, the researcher provides pre-defined dictionaries (lists) of words associated with a given emotion, such as negativity. The target text is then deconstructed into individual words (or tokens) and the frequencies of words contained in a given dictionary are then calculated. For example, Loughran and McDonald (2011) construct a dictionary of words that express negativity in financial contexts, which they then use to measure the negativity in company 10-K filings and relate that negativity to financial outcomes.

We apply the Kanjoya predictive models to a large set of economic and financial articles from 16 major U.S. newspapers dating back to 1980. We construct a monthly time series index, for each of several emotions, by estimating sample-month fixed effects from a regression of article scores on fixed effects for sample-month and newspaper*article-type, where article type is either regular article or editorial. The newspaper*article-type fixed effects control for changes in newspaper coverage over the sample and systematic differences in tone across newspapers and article types.

Kanjoya provides a predictive model for a large set of emotions. To keep this study of manageable scope, we focus our analysis on Kanjoya's measure of emotional valence—what

[1] The Kanjoya entity is now a subsidiary of Ultimate Software.

[2] See Socher, Pennington, Huang, Ng, and Manning (2011), Potts (2010), and Sudhof, Goméz Emilsson, Maas, and Potts (2014).

is referred to throughout the paper as "negativity." The psychology literature demonstrates that any emotion can be parsimoniously defined by assigning it a score along two primary dimensions: a valence (i.e., positive or negative) and arousal (i.e., activation and deactivation)(see Tellegen (1985), Lang, Bradley, and Cuthbert (1990), Bradley, Greenwald, Petry, and Lang (1992), and Posner, Russell, and Peterson (2005)). For comparison purposes, we also examine three additional measures: "worried," "satisfied," and a lexical "negativity" measure. The "worried" measure presumably includes a forward-looking component, and may align better with the informational view of sentiment, while the "satisfied" measure provides a measure in the positive valence range and may capture different aspects of emotion than the pure valence measure. The lexical measure can be thought of as a type of hybrid model insofar as Kanjoya uses a structured learning algorithm to assign words to a negativity dictionary.

We examine how well each of these four news sentiment measures correlate with current business conditions, the degree to which innovations in these measures drive future economic activity, and how much they aid in terms of predictive accuracy. Overall, we find that the news sentiment indexes correlate strongly with contemporaneous key business cycle indicators. Furthermore, we find that innovations to news sentiment orthogonal to business cycle fundamentals, or what can be referred to as "news sentiment shocks," predict future economic activity. Generally, we find that economic activity responds to shocks to the model-based measures in a stronger fashion than to the lexical measure studied in this paper. Furthermore, in most cases, the news sentiment measures outperform the University of Michigan and Conference board measures in head-to-head comparisons. In particular, the news sentiment measures have more predictive information in terms of the federal funds rate, consumption, employment, inflation, industrial production, and the S&P500. For some of the economic outcomes, there is evidence that the news sentiment measures have significant predictive power even after conditioning on these survey-based measures.

2 Approaches to Measuring Sentiment

2.1 Surveys

The traditional approach to measuring economic sentiment is to construct indexes based on surveys. Two prominent examples are the Michigan Consumer Sentiment index and the Conference Board's Consumer Confidence index.[3] In Section 6, we compare the explanatory power for forecasting macroeconomic outcomes of these indexes compared with that of the news sentiment index we construct in this paper. Here we provide some background on these survey-based measures of sentiment before describing the computational text analysis approach.

2.1.1 Michigan Consumer Sentiment

The University of Michigan's Consumer Sentiment (CS) Index dates back to the late 1940s.[4] It is based on a monthly telephone survey of at least 500 respondents across the U.S. The index is constructed as a normalized sum of the relative scores (percent of responses that are favorable minus percent of responses that are unfavorable) from the following five questions:

(1) "We are interested in how people are getting along financially these days. Would you say that you (and your family living there) are better off or worse off financially than you were a year ago?"

(2) "Now looking ahead–do you think that a year from now you (and your family living there) will be better off financially, or worse off, or just about the same as now?"

(3) "Now turning to business conditions in the country as a whole–do you think that during the next twelve months we'll have good times financially, or bad times, or what?"

(4) "Looking ahead, which would you say is more likely–that in the country as a whole we'll have continuous good times during the next five years or so, or that we will have periods of widespread unemployment or depression, or what?"

(5) "About the big things people buy for their homes–such as furniture, a refrigerator,

[3] Such survey-based indexes are not limited to consumer surveys. There are also sentiment/confidence surveys of business decision-makers such as the surveys underlying the Conference Board's "CEO Confidence Index" or the National Federation of Independent Businesses' "Small Business Optimism Index".

[4] Further details about the Michigan Consumer Sentiment index and the underlying survey can be found at: `https://data.sca.isr.umich.edu`.

stove, television, and things like that. Generally speaking, do you think now is a good or bad time for people to buy major household items?"

2.1.2 Conference Board Consumer Confidence

The Conference Board's Consumer Confidence (CC) index dates back to 1967 and is based on their Consumer Confidence Survey. Since 1977, this survey has been conducted monthly. The Conference Board aims to get responses from roughly 3,000 households. Similar to the Michigan CS index, the CC index is based on the responses to five survey questions. The first two questions relate to current business conditions and current employment conditions. The other three questions relate to expectations of business conditions, employment conditions, and family income six months ahead.

2.2 Sentiment Text Analysis

Sentiment analysis utilizes tools from the field of computational linguistics to quantify the emotional content in a set of text.[5] The sentiment of text (i.e., a word, phrase, sentence, etc.) typically is characterized first by its emotivity (the degree to which the text is emotive versus neutral/objective) and then by the degrees to which it exhibits certain emotions. Drawing from the psychology literature, emotions are typically characterized by the following two dimensions (at least): valence and arousal. Valence captures how positive or negative is the emotion. Arousal measures how charged is the emotion. (Emotions are sometimes also charactered on a temporal dimension, reflecting whether the emotion tends to express feelings that are forward-looking, backward-looking, or present-looking).

Figure 1 provides an example, from Bradley and Lang (1999)'s survey evidence, of this two-dimensional characterization for a common set of emotions. For each indicated emotion, the figure indicates the relative level of valence (on the x axis) – from maximal negativity to maximal positivity – and the relative level of arousal (on the y axis) associated with that emotion. An emotion such as satisfied is high on the valence scale and roughly neutral on the arousal scale. Excited is similar in terms of valence but much higher in arousal. Stress

[5]Sentiment analysis is one type of text analysis. Another type that is increasingly being used in economics and finance is topic analysis (see, for example, Hansen and McMahon (2016) and Hansen, McMahon, and Prat (2014)). Topic analysis identifies the topics most often discussed in a body of text and seeks to gather insights into how and why certain topics are discussed at various times by various speakers.

is similar to Excited in arousal but on the other end of the spectrum of valence, i.e., high in negativity.

There are two general approaches to quantifying sentiment in text. The first is known as the "Bag of Words" or **Lexical** approach. It relies on a pre-defined dictionary (or set of dictionaries, one for each emotion) of words that are associated with that emotion. There are a number of examples of applications of this approach in Economics/Finance. Loughran and McDonald (2011) construct their own dictionary of negativity, arguing that Harvard Psychosociological Dictionary considers as negative many words that are neutral in a financial/economic context (like tax, costs, capital, expense, liability, risk, excess, depreciation). Heston and Sinha (2015) measure valence in news articles about companies and estimate their impact on those companies' stock returns. They use the Harvard Psychosociological Dictionary along with Loughran and McDonald's dictionary. Fraiberger (2016) measures valence in international news articles using dictionaries from Loughran and McDonald (2011) and Young and Soroka (2012), and uses these measures to improve GDP forecasts.[6]

The second approach uses Natural Language Processing (NLP) tools, which is a nascent subfield of computational linguistics and relies on machine learning techniques.[7] NLP sentiment analysis attempts to extract emotional content from a set of text based both on word choice (lexicon) and the context (combinations and structure) of words. It estimates the emotional content of text using a predictive model that is trained on a large corpus of text containing a mapping between utterances and emotions. For instance, Kanjoya's model is trained using data from their "Experience Project" social network website, where millions of users report their current emotion(s) while simultaneously posting text describing recent experiences, issues, and feelings. This data set, containing *both* emotion and text, allows for the application of structured machine learning techniques for building a high-quality predictive model of emotions. The ability to train a predictive model using structured machine learning, rather than rely on a researcher's own subjective and limited identification of terms associated with each emotion, is an important potential advantage of the NLP, or **model-based** approach over the lexical approach. One the other hand, the model approach is only as good as its training set (i.e., the data set pairing text and emotions) and so if the text

[6]The study of economic policy uncertainty by Baker, Bloom, and Davis (2015) also uses a lexical-based measure as part of its uncertainty index. That measure is a simple count of news articles containing terms such as "uncertain" and "not certain" along with terms related to economic policy.

[7]See Liu (2010) for a detailed description of the NLP approach to sentiment analysis.

in the training set is not representative of the text to which the predictive model is being applied, the predictions may well be inaccurate.

One can also combine the lexical and model approaches. For instance, a hybrid approach is to use similar machine learning techniques and the same corpus of text used in NLP approach to construct richer and more scientific dictionaries for a given emotion. In particular, Kanjoya's pairings of self-identified emotions with text allows one to construct a dictionary of words that tend to be associated with any given emotion, such as negativity, sadness, anger, optimism, etc.

In our application below, we construct indexes of sentiment both using the NLP/model-based approach and using emotional scores obtained from the hybrid approach (i.e., scores based on a dictionary for that emotion containing words and weights identified from a machine learning algorithm).

3 Data

We construct a time series index of economic news sentiment. The construction of the index consists of three broad steps: (1) obtaining a large corpus of economic news articles from LexisNexis, (2) applying sentiment analysis algorithms to each article to score the article in terms of several different sentiment measures, and (3) constructing an aggregate monthly time series for each sentiment measure. We describe each step in the following three subsections. In a subsequent subsection, we describe the economic variables which we empirically relate to the news index in the remainder of the paper.

3.1 A Corpus of Economic News Articles

We purchased a large archive of newspaper articles from the news aggregator service, LexisNexis (LN). We pulled all newspaper articles (including editorials) from 1980 to present from 16 major U.S. newspapers[8] satisfying the following criteria:

[8]The newspapers are: Atlanta Journal-Constitution, Boston Globe, Chicago Tribune, Detroit Free Press, Houston Chronicle, Los Angelos Times, Memphis Commercial Appeal, Miami Herald, Minneapolis Star Tribune, New Orleans Times-Picayune, New York Times, Philadelphia Inquirer, San Francisco Chronicle, Seattle Times, St. Louis Post-Dispatch, and The Washington Post.

(1) LN classified “country subject” as “United States” (with an LN “relevance” threshold of at least 85%).
(2) LN classified “topic subject” as “Economy” or “Economic” (with an LN “relevance” threshold of at least 85%).
(3) LN did NOT classify the article as a “Brief” or “Summary” or “Digest.“
(4) Article had 200 words or longer.
(5) Article contained at least one of the following words: said, says, told, stated, wrote, reported.

Restrictions 1 and 2 allow us to focus on articles that are related to U.S. economic news. Restriction 3 mitigates duplication of news in that those articles are typically summaries of articles appearing elsewhere. Restriction 4 is useful because it is more difficult for the sentiment analysis algorithm to accurately score the emotions of an article when the article has very few words. Restriction 5 allows us to focus on articles most likely to express sentiment – i.e., to have sufficient emotivity. Articles containing those words typically express the sentiment (often, but not always, a quotation) of a person or group of people. After imposing these criteria, our data pull yielded approximately 231,000 articles.

3.2 Sentiment Scores of News Articles

Our objective in this step is to construct models that can accurately summarize the affective content of newspaper articles and editorials. Since our models must be sensitive to text spanning a wide range of topics, ideological bents, geographic regions, and time periods, it is imperative the models are trained on a large, robust corpus of expressive utterances that is not specific to any one topical domain such as, for example, financial reports.

Ideally, one would identify the emotional content of a given article by crowdsourcing every article, asking a large number of people to score the article in terms of the degree to which the article expresses a given emotion. However, for large corpuses of text such as our set of 231,000 news articles, crowdsourcing is infeasible, necessitating a computational approach.

We utilize a proprietary predictive sentiment model provided by the company, Kanjoya. As mentioned in the previous section, Kanjoya's model is trained using data from their “Experience Project" social network website. The Experience Project corpus spans millions of distinct emotional utterances with granular emotional labels (over 130 total) and no given topical focus. Samples of this corpus have been previously shown to yield high quality sentiment and emotion models.[9] The corpus was collected over 8 years and comprises contributions from a population that is highly diverse in terms of gender, age, race/ethnicity, and geographical location. Crucially, labels are assigned to textual utterances based on the author?s own stated corresponding emotion or the reactions of other users when interacting with the content. Observing both these emotion labels (outputs) and the text (inputs) assigned that label allows for structured machine learning techniques to train a predictive model.

We follow prior work using the Experience Project data in terms of text processing and model-training. However, we are able to utilize the entire database of Experience Project utterances, whereas prior work relied on smaller research samples.

As mentioned above, we use this corpus both to construct, for a given emotion, both a lexical dictionary and a predictive model that can be used to measure the degree to which a piece of text exhibits that emotion. We first automatically construct lexicons that relate emotions to words and phrases that are indicative of the presence or absence of the emotion in text. We transform documents into lists of constituent words and phrases by tokenizing and part-of-speech tagging text. To extract phrases, we rely on a hand-built set of high-precision templates to identify meaningful textual units. The approach means that we can work with coherent phrases that are not artificially bounded in size. [10] Next, we use vector space models (VSMs) to describe the frequency of usage of each word or phrase for each emotional label in the corpus. Finally, we use the G-test (a log-odds version of the chi-squared test) to derive lexicon scores from the count distributions of the words and phrases. The hypothesis is that the word or phrase's distribution is independent of the emotional labels; the G-test is assessing our confidence that the observed values differ from our expectations under this null hypothesis. This approach is able to incorporate the entire corpus and full set of emotional labels. The resulting lexicon can be used to identify words or phrases with high mutual

[9]See Socher, Pennington, Huang, Ng, and Manning (2011), Potts (2010), and Sudhof, Goméz Emilsson, Maas, and Potts (2014).

[10]In contrast, n-gram based models are constrained by their chosen n, they are much more susceptible to identifying semantically incoherent blocks of text, and they typically offer no meaningful performance gains over part-of-speech tagging.

information content for any emotion or sentiment category.

Given such lexicons, assigning a lexical score to a newspaper article or editorial is simple. We extract words and phrases from the newspaper text in the same manner as above, and, for each emotion or sentiment category of interest, we sum the lexicon scores for each word or phrase contained in both the newspaper text and the lexicon. We then divide by the total length of the document to arrive at a normalized score of a certain emotional category in the text. The lexical score we use in this study is the within-article difference between the negative and positive lexical score—this controls for the fact some articles may be more "emotional" and have a high level of both negative and positive words.

Table 1 shows the most common words from our newspaper text that correspond to various emotions according to our lexicon. For example, we find that the most common "negative" words appearing in our sample of newspaper articles are: wrong, problem, difficult, weak, worst, disturbing, concerned, terrible, disappointing, and bad.

Lexicons are powerful tools for identifying the words and phrases carrying affective weight in text, but they are not able to incorporate conflicting signals in order to make probabilistic decisions about text. To build such a model, we use the corpus of annotated Experience Project utterances as a training set to train hierarchical log-linear classifiers that can estimate sentiment and emotion. The previously generated lexicon is used to extract features from text, and word- and phrase-level features are combined with additional syntactical features as inputs to the training algorithm. Given a newspaper text, the resulting models compute probability scores for each sentiment or emotion category of interest. Probability scores range from 0 (not present) to 1 (present) and provide a measure of the model's confidence that a given emotion is present in the text provided.

We provide two examples of Kanjoya's scoring of newspaper articles in Appendix A. The first article is from the Chicago Tribune and was published on July 25, 2012. The headline of the article is "Analysis: Fewer to Get Health Insurance," and discusses a report by the Congressional Budget Office concluding that fewer individuals would get health insurance as a result of a Supreme Court decision allowing states to opt out of part of the Affordable Care Act. The model-based method assigns this article a very high negativity score (near the 99th percentile of articles) while the lexical-based method assigns a low negativity score (28th percentile).[11]

[11] The article also is assigned a very low satisfied score (2nd percentile), consistent with the high negativity

The second article is from the New Orleans Times-Picayune on April 28, 1999. The headline is "U.S. Income Growth Robust in 1998" and reports on newly released data from the BEA on income growth by state in 1998. The article discusses the rapid and widespread growth in real incomes in that year and mentions other positive economic news such as the facts that "unemployment is near a 29-year low" and "brisk spending by Americans has been largely responsible for keeping the U.S. economy growing while other countries have been hurt by a global economic downturn...." This article receives a low negativity-model score (3rd percentile of articles), but a high negativity-lexical score (99th percentile).[12] It appears that the lexical-based method is misled, in a sense, by the usage of words like "low," "hurt," and "downturn," while the model-based method is better able to understand the overall positive context of these phrases.

3.3 Construction of Monthly News Sentiment Index

Our final objective is to construct a monthly index of news sentiment in order to assess whether and how news sentiment affects, and is affected by, economic outcomes. We construct this index by estimating the month fixed effects ($\hat{f}^i_{t(a)}$) from the following regression over articles (indexed by a):

$$s^i_a = f^i_{t(a)} + f^i_{p(a),j(a)} + \varepsilon^i_a. \tag{1}$$

where s^i_a is the score for sentiment $i \in \{\text{negativity-model, negativity-lexical, satisfied, worried}\}$ for article a and $f^i_{t(a)}$ is a sample-month (t) fixed effect. Newspapers are indexed by j and article type – either editorial or regular article – is indexed by p. $f^i_{p(a),j(a)}$ is thus a newspaper*type fixed effect.

Allowing for newspaper*type fixed effects ensures that the index is independent of changes over time in the composition of the sample across newspapers and editorials versus regular articles. This can be important because the tone of articles differ considerably across newspapers and between editorials and regular articles within a newspaper. This can be seen by looking at the estimated newspaper*type fixed effects, as shown in Figure 2. The top panel shows the results for the negativity-lexical scores; the bottom panel shows the results for the

model score.

[12]The article also is assigned a low satisfied score (19th percentile).

negativity-model scores. Note that the omitted category in both cases is regular articles in the Atlanta Journal-Constitution. The scores are normalized by their standard deviation, so the x-axis values are in units of one standard deviation.

Sentiment scores vary substantially across newspapers and type. For instance, on average a regular article in the New Orleans Times-Picayune is about 1.4 standard deviations more positive (according to Negativity-Model) than an article in the Atlanta Journal-Constitution. We also find substantial variation in average sentiment across editorials, with the Atlanta Journal-Constitution being the most negative and the Seattle Times being the most positive (according to negativity-model).

4 Descriptive Analysis

In this section, we present some descriptive analysis of the monthly sentiment indexes. Recall the monthly sentiment indexes are constructed from the estimated month fixed effects from regressing sentiment scores by article on month and newspaper*type fixed effects (see equation (1)): $s_{i,p,j} = \hat{f}^{i}_{p(a),j(a)}$, where i is a particular measure of sentiment (e.g., negativity-model), p is a type of article, and j a specific newspaper.

Figure 3 plots our primary two measures of news sentiment over time. The Negativity-Lexical series is colored blue; the negativity-model series is colored orange. First, notice the two series are strongly correlated. Second, both tend to spike during months of key historical events affecting economic outcomes and financial markets, such as the start of the first Gulf War, the Russian financial crisis, Lehman Brothers bankruptcy, and when S&P lowered its credit rating on U.S. government debt in August 2011.

Table 2 shows the correlations among all four of our measures of sentiment. As one would expect, negativity-model, negativity-lexical, and worried are all positively correlated with each other and are all negatively correlated with Satisfied. The negative correlation with satisfied is especially large for Negativity-Model.

Lastly, we assess how these four measures of news sentiment are correlated (contemporaneously) with key economic outcomes. The outcomes we consider are the month-over-month change in the fed funds rate (FFR) and log changes in the S&P 500 stock price index, real personal consumption expenditures (PCE), total nonfarm employment, industrial production (IP), and the PCE price index (PCEPI).

The correlations are shown in Figure 4. Cells are color-coded with positive values colored green and negative values colored red. Negativity-model, negativity-lexical, and worried are each negatively correlated with all five of the real-side economic outcomes (i.e., the outcomes other than PCEPI inflation). Satisfied is positively correlated with all five real-side outcomes. The pattern of correlations for PCEPI inflation is less clear: Satisfied is negatively correlated with inflation, as is negativity-lexical and worried, while negativity-model has a very small positive correlation with inflation.

Although these are only simple correlations, the fact that when real economic activity is high, negativity and worried are low and satisfied is high, suggests that these news sentiment indices are not simply noise. Rather, they appear to either reflect or affect economic activity. We turn next to assessing the timing and direction of causality.

5 Estimating the Effects of News Sentiment Shocks on Economic Activity

To assess whether news sentiment affects economic activity, we perform an impulse response analysis using the local projection method of Jorda (2005), which is similar to but less restrictive than the estimated impulse response function from a vector auto-regression. The news sentiment shock is constructed as the component of the news sentiment series that is orthogonal to current and 12 lags of economic activity as well as 12 lags of itself. That is for each forecast horizon h, a distinct regression is run of the economic variable on current news sentiment as well as lagged values the news sentiment index and economic variables:

$$y_{j,t+h} = \beta_i^h S_{i,t} + \sum_{l=1}^{12} \alpha_k S_{i,t-l} + A \sum_{l=0}^{12} \mathbf{Y}_{t-l} + \varepsilon_{i,t}. \tag{2}$$

where the vector $\mathbf{Y}$ includes the federal funds rate (FFR), and the logarithms of real consumption (C), payroll employment (E), the personal consumption expenditures price index (PCEPI), and the S&P 500 index (SP500). These are the same economic variables used in Baker, Bloom, and Davis (2015) and were chosen because they are available monthly and cover broad aspects of the economy. The impulse response function from a news sentiment shock on economic variable y are the estimates of $\hat{\beta}_{i,t}^h$ from equation (2). We consider horizons from $h = 0$ to 12 months. We focus on four measures of news sentiment, S_i, (1) negativity-model, (2) negativity-lexical, (3) worried , and (4) satisfied. The impulse responses are shown in Figure 5 along with 90 percent confidence bands. All shocks are normalized to one standard deviation.

5.1 Results

Most variables respond to the news shocks as they would a typical demand shock. For instance, a negative sentiment shock causes a smooth decline in the federal funds rate, employment, and inflation, while we see a reverse for the the satisfied shock—a shock that represents positive news surprises. Specifically, a satisfied shock tends to raise employment, industrial production, federal funds rate, and to some extent, consumption. The effects take some time to unfold—a statistically significant effect on the PCEPI does not merge until around 3 months after the shock, while an impact on FFR does not emerge until about 6 months after the shock. We see no discernable patterns for the S&P 500 index. Overall, it appears that there are responses to news sentiment shocks. A negative news sentiment shock causes declines in employment, industrial production, the PCEPI, and the federal funds rate.

We gain some insight on the relative performance of the model-based and lexical-based measures by comparing the impulse responses of the two types of negativity shocks. In most cases, the model-based measure of sentiment produces smoother, more statistically significant impulse response functions. This is particularly true in terms of the FFR and PCEPI, where a one standard deviation negativity-model shock causes a 10 basis point reduction in the FFR and a 4 basis point decline in the PCEPI, but much smaller effects from the negativity-lexical measure. Similarly, the negativity-model shock causes a smooth reduction in employment, albeit not statistically significant, however, the lexical-based measure produces no discernible effect at all.

6 Predictive Accuracy of News Sentiment

The previous exercise examined the direction and size of the impact of news sentiment on economic activity. Using the local-projection method as in the previous section, we examine whether our news sentiment measures contain any predictive information about future economic conditions. In the following analysis, we perform both in-sample and pseudo-out-of-sample forecasting comparison tests. In-sample tests are advantageous because they use all of the available data, however, recent studies have discussed reasons why out-of-sample tests may have better finite-sample performance (Hansen (2010), Diebold (2015)).

As a first exercise, we compare 12-month ahead forecasts of the baseline model and the baseline model including various news sentiment measures. We also include a specification with all 14 of Kanjoya's sentiment measures, as well as the first three principal components of all of Kanjoya's sentiment measures.[13] Specifically, Table 1 reports forecasting test results comparing the following baseline model and an extended model including sentiment at a 12-month forecasting horizon:

$$\textbf{Baseline:} \quad y_{i,t+12} = A\sum_{l=0}^{12}\mathbf{Y}_{t-l} + \varepsilon_{i,t} \tag{3}$$

$$\textbf{Baseline + News Sentiment:} \quad y_{i,t+12} = A\sum_{l=0}^{12}\mathbf{Y}_{t-l} + \sum_{l=1}^{12}\alpha_k S_{i,t-l} + \varepsilon_{i,t} \tag{4}$$

where y represents the economic variable of interest j, Y represents the full vector of economic variables and their 12 lags, and s_i represents a particular sentiment index i. In addition to the four individual sentiment index measures assessed above (negativity-model, negativity-lexical, worried, and satisfied), we also include a specification where we include the first four principal components (making up 50% of the total variation) of all sentiment measures, as well a specification which includes all four sentiment measures. To compare the forecasts with and without the news sentiment scores, we report four statistics: (1) the p-value of the F-statistic testing the joint hypothesis that the coefficients on the sentiment measure and its lags equal zero, (2) the difference between the Bayesian Information Criterion (BIC) of the baseline and sentiment in-sample models, (3) the analogous difference in the Akaike Information Criterion (AIC) of the in-sample models, and (4) the difference in the Giacomini-White statistic (GW) between the baseline and sentiment pseudo-out-of-sample forecasts, along with statistical significance. The pseudo-out-of-sample forecasts are taken using 15-year rolling window estimates. Note that a decline in the AIC, BIC, or GW statistic implies

[13] Kanjoya's sentiment measures include: angry, sad, upset, happy, optimistic, confused, thoughtful, confident, stressed, satisfied, worried, annoyed, appreciative, and excited.

that the inclusion of sentiment improves the forecast fit.

6.1 Results

Results from Table 3 indicate that model with sentiment improves the pseudo out-of-sample forecast across all news sentiment measures for all of the economic measures used in this study. Specifically, the GW statistic declines with the inclusion of new sentiment in all cases, and most cases the decline is statistically significant.[14] In terms of in-sample tests, the results are not as clear cut—including news sentiment improves the forecast for some economic variables and some sentiment measures but not others. We see a statistically significant F-test and decline in the AIC across a broad range of sentiment measures for the PCE price index, and to some extent for the FFR. For the other economic variables there does not seem to be an advantage in the in-sample forecast over and above the baseline model.

We also perform two additional exercises which assess whether the news sentiment measures provide any new information relative to measures of sentiment that currently are available to the public: the Michigan consumer sentiment index (MCS) and the Conference Board's consumer confidence index (CBCC). The first test is similar to the exercise above, nesting the baseline model into that of the model with news sentiment:

$$\textbf{Extended Baseline:} \quad y_{i,t+12} = A\sum_{l=0}^{12}\mathbf{Y}_{t-l} + B\sum_{l=0}^{12}\mathbf{Z}_{t-l} + \varepsilon_{i,t} \tag{5}$$

$$\textbf{Extended Baseline + News Sentiment:} \quad y_{i,t+12} = A\sum_{l=0}^{12}\mathbf{Y}_{t-l} + B\sum_{l=0}^{12}\mathbf{Z}_{t-l} + \sum_{l=1}^{12}\alpha_k S_{i,t} + \varepsilon_{i,t} \tag{6}$$

where the vector Z includes 12 lags of both the MCS and the CBCC. This exercise tests whether the news sentiment measures provide any additional forecasting information over and above the MCS and CBCC.

The results of this exercise are shown in 4. Similar to the results in 3, the GW statistic declines for each economic measure in every news sentiment measure, indicating that the

[14]Significance stars are shown in the table. */**/*** indicate significance at the 10/5/1 percent level.

news sentiment measures are improving the out-of-sample forecasting performance of the model in a robust manner. Also similar to 3, the news sentiment measures improve the in-sample forecasting performance for the PCE price index across all news sentiment indexes, but the results are less robust for the other economic variables. It is interesting to note that the principal components of the news sentiment measures improves the in-sample forecasts of the federal funds rate, real consumption, industrial production, and employment according to the F-statistic and the AIC.

Our second test is a non-nested exercise where we compare the performance of the news sentiment with that of the publicly available measures (Z)—a standard horse-race comparison:

$$\textbf{Extended Baseline:}\quad y_{i,t+12} = A\sum_{l=0}^{12}\mathbf{Y}_{t-l} + B\sum_{l=0}^{12}\mathbf{Z}_{t-l} + \varepsilon_{i,t} \tag{7}$$

$$\textbf{Baseline + News Sentiment:}\quad y_{i,t+12} = A\sum_{l=0}^{12}\mathbf{Y}_{t-l} + \sum_{l=1}^{12}\alpha_k S_{i,t} + \varepsilon_{i,t} \tag{8}$$

This exercise assesses whether the news sentiment of the publicly available measures are providing more predictive information. The results from this exercise are shown in 5. The out-of-sample forecasts are either lower or statistically insignificant (at the 5 percent level) using the the news sentiment measures. Interestingly, the in-sample results indicate that in most cases the news sentiment have a better in-sample forecasting fit in terms of the AIC and BIC. The contrast with the previous results is due to the fact that both of these tests include a degrees-of-freedom adjustment, which is larger in the nested model than the horse-race model. In other words, the "Extended + News Sentiment" model is heavily penalized in the AIC and BIC formula for including many variables. Overall, in most cases the news sentiment measures have more predictive information both in-sample and out-of-sample, than the publicly available sentiment measures in a head-to-head comparison. The nested models indicate that the news sentiment measures improve out-of-sample forecasting performance relative to the baseline, meaning there is sufficient orthogonal information in the news sentiment measures to aid in out-of-sample forecasting.

7 Conclusion

In this paper, we developed a measure of news sentiment using computational text analysis of economic and financial newspaper articles. We assessed how well this index correlates with and predicts economic outcomes. Our results corroborate previous studies showing that sentiment has predictive power on future economic activity. Specifically, we have shown that sentiment extracted from newspaper articles correlates with both contemporaneous and future key business cycle indicators. In a head-to-head comparison, these news sentiment measures perform better than both the University of Michigan and Conference board measures of consumer sentiment. In particular, the news sentiment measures have more predictive information in terms of the federal funds rate, consumption, employment, inflation, industrial production, and the S&P500. For some of the economic outcomes, there is evidence that the news sentiment measures have significant predictive power even after conditioning on these survey-based measures.

We conclude that these methods of sentiment text analysis hold great promise for improving our understanding of news sentiment shocks and how they affect the economy.

References

ANGELETOS, G.-M., AND J. LA'O (2013): "Sentiments," *Econometrica*, 81(2), 739–779.

BAKER, S. R., N. BLOOM, AND S. J. DAVIS (2015): "Measuring economic policy uncertainty," (64986).

BARSKY, R. B., AND E. R. SIMS (2012): "Information, animal spirits, and the meaning of innovations in consumer confidence," *The American Economic Review*, 102(4), 1343–1377.

BENHABIB, J., P. WANG, AND Y. WEN (2015): "Sentiments and aggregate demand fluctuations," *Econometrica*, 83(2), 549–585.

BLANCHARD, O. (1993): "Consumption and the Recession of 1990-1991," *The American Economic Review*, 83(2), 270–274.

BRADLEY, M. M., M. K. GREENWALD, M. C. PETRY, AND P. J. LANG (1992): "Remembering pictures: pleasure and arousal in memory.," *Journal of experimental psychology: Learning, Memory, and Cognition*, 18(2), 379.

BRADLEY, M. M., AND P. J. LANG (1999): "Affective Norms for English Words (ANEW): Instruction Manual and Affective Ratings," Technical Report C-1, The Center for Research in Psychophysiology, University of Florida.

BRAM, J., AND S. C. LUDVIGSON (1998): "Does consumer confidence forecast household expenditure? A sentiment index horse race," *Economic Policy Review*, 4(2).

CARROLL, C. D., J. C. FUHRER, AND D. W. WILCOX (1994): "Does consumer sentiment forecast household spending? If so, why?," *The American Economic Review*, 84(5), 1397–1408.

DIEBOLD, F. X. (2015): "Comparing predictive accuracy, twenty years later: A personal perspective on the use and abuse of Diebold–Mariano tests," *Journal of Business & Economic Statistics*, 33(1), 1–1.

Fraiberger, S. P. (2016): "News Sentiment and Cross-Country Fluctuations," *Available at SSRN.*

Hansen, P. R. (2010): "A winners curse for econometric models: on the joint distribution of in-sample fit and out-of-sample fit and its implications for model selection," *Research Paper*, pp. 1–39.

Hansen, S., and M. McMahon (2016): "Shocking language: Understanding the macroeconomic effects of central bank communication," *Journal of International Economics*, 99, S114–S133.

Hansen, S., M. McMahon, and A. Prat (2014): "Transparency and Deliberation within the FOMC: a Computational Linguistics Approach," (1411).

Heston, S. L., and N. R. Sinha (2015): "News versus sentiment: Predicting stock returns from news stories," *Robert H. Smith School Research Paper.*

Jorda, O. (2005): "Estimation and Inference of Impulse Responses by Local Projections," *American Economic Review*, 95(1), 161–182.

Lang, P. J., M. M. Bradley, and B. N. Cuthbert (1990): "Emotion, attention, and the startle reflex.," *Psychological review*, 97(3), 377.

Liu, B. (2010): "Sentiment Analysis and Subjectivity.," *Handbook of natural language processing*, 2, 627–666.

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

Ludvigson, S. C. (2004): "Consumer confidence and consumer spending," *The Journal of Economic Perspectives*, 18(2), 29–50.

Posner, J., J. A. Russell, and B. S. Peterson (2005): "The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology," *Development and psychopathology*, 17(03), 715–734.

Potts, C. (2010): "On the negativity of negation," in *Semantics and Linguistic Theory*, vol. 20, pp. 636–659.

Socher, R., J. Pennington, E. H. Huang, A. Y. Ng, and C. D. Manning (2011): "Semi-supervised recursive autoencoders for predicting sentiment distributions," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pp. 151–161. Association for Computational Linguistics.

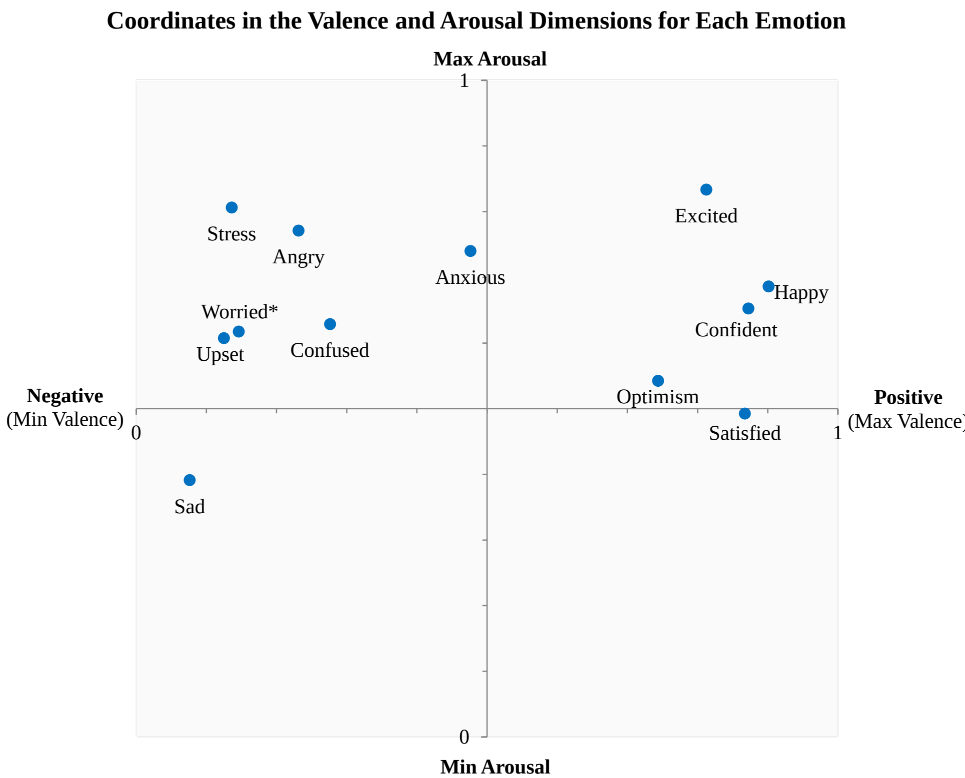
Souleles, N. S. (2004): "Expectations, heterogeneous forecast errors, and consumption: Micro evidence from the Michigan consumer sentiment surveys," *Journal of Money, Credit, and Banking*, 36(1), 39–72.

Sudhof, M., A. Goméz Emilsson, A. L. Maas, and C. Potts (2014): "Sentiment expression conditioned by affective transitions and social forces," in *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 1136–1145. ACM.

Tellegen, A. (1985): "Structures of mood and personality and their relevance to assessing anxiety, with an emphasis on self-report.," .

Young, L., and S. Soroka (2012): "Affective news: The automated coding of sentiment in political texts," *Political Communication*, 29(2), 205–231.

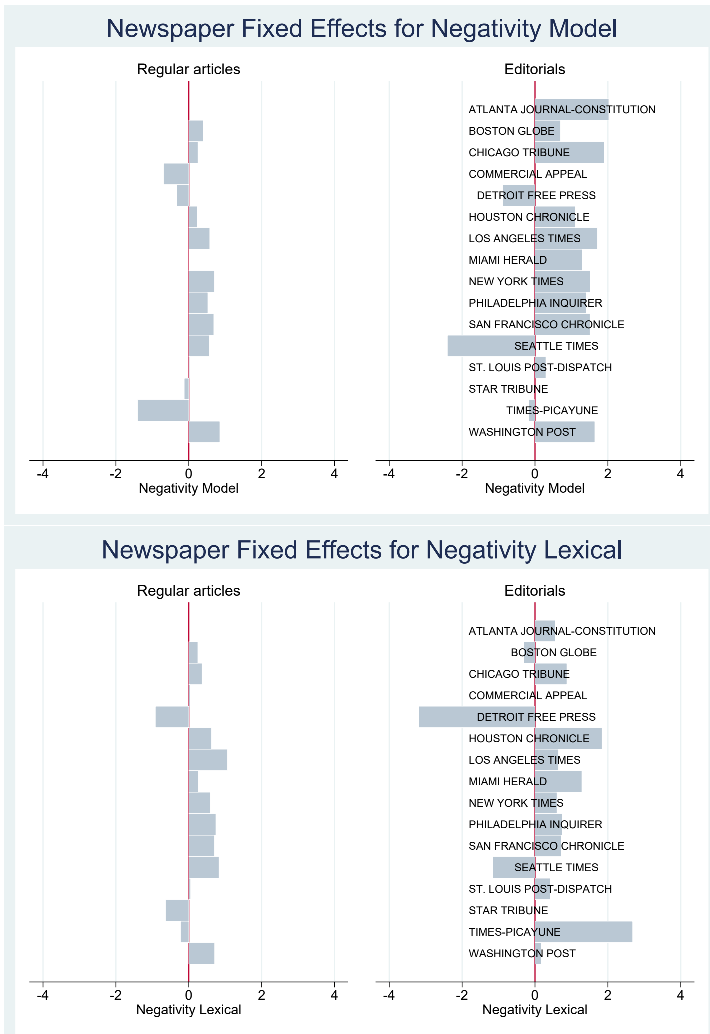
Figure 1: Characterization of Emotions Along Valence and Arousal Dimensions



Source: Survey-participant means reported in Bradley & Lang (1999). *"Worried" is a synonym of "troubled"; Bradley & Lang include "troubled" instead of "worried."

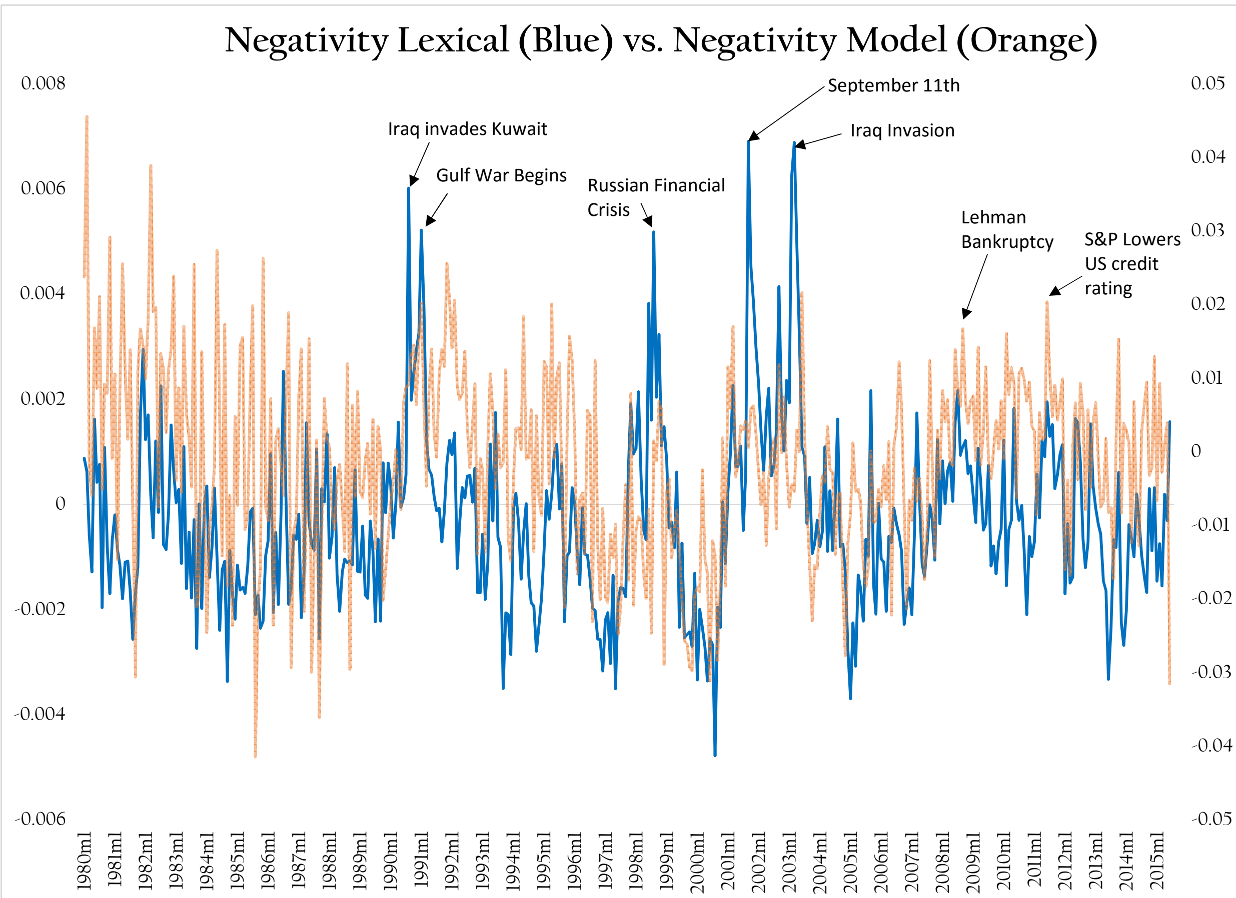
Figure 2: Newspaper-by-Type Fixed Effects

Figure 1: Newspaper-by-Type Fixed Effects



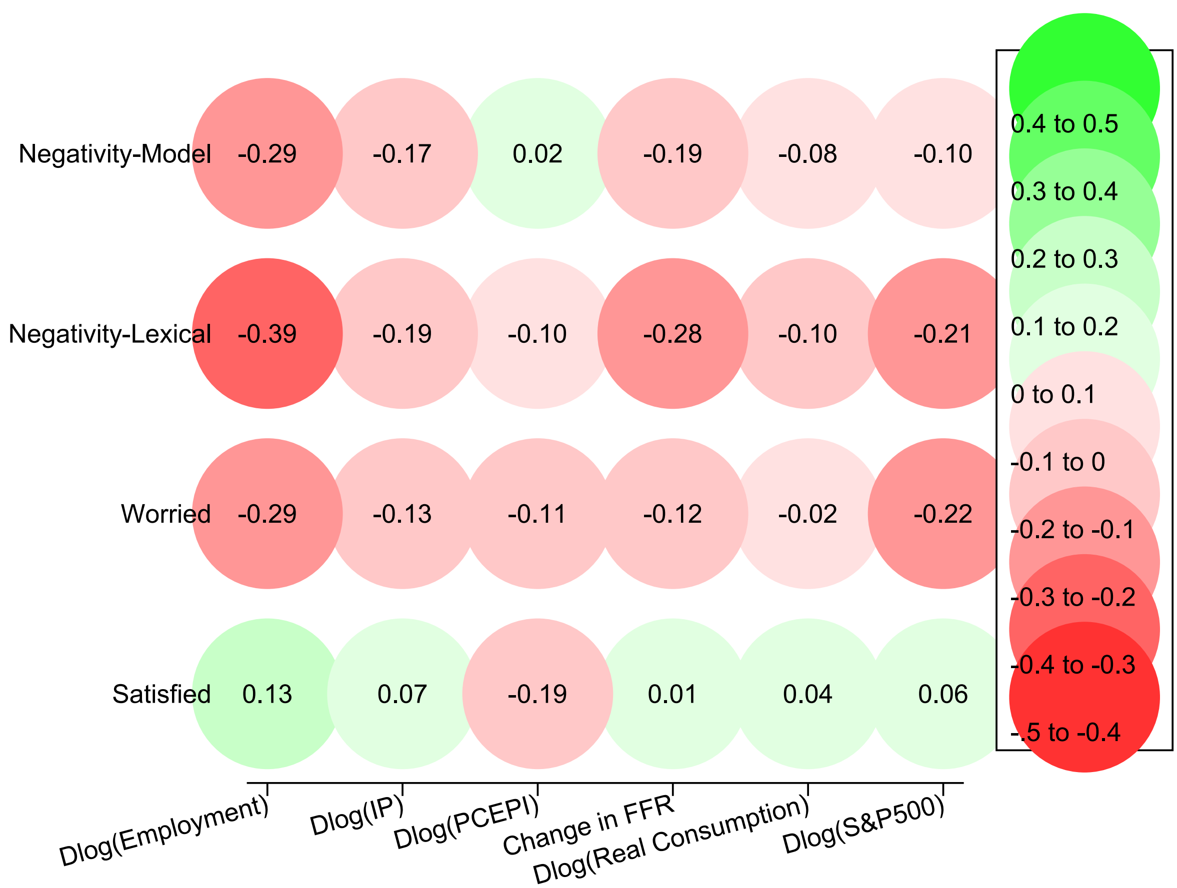
Notes: Shown are the normalized point estimates of the newspaper-type fixed effects for the negativity-model (top panel) and negativity-lexical (bottom panel) for the 16 newspapers in our sample, where type indicates either news or editorial article. A separate regression is run for each sentiment measure, which also includes time-dummies (in months).

Figure 3: Sentiment Indexes Over Time



Notes: Shown are the point estimates of the time dummies (in months) for the negativity-model (orange line) and negativity-lexical (blue line). A separate regression is run for each sentiment measure, which also includes newspaper-type dummies.

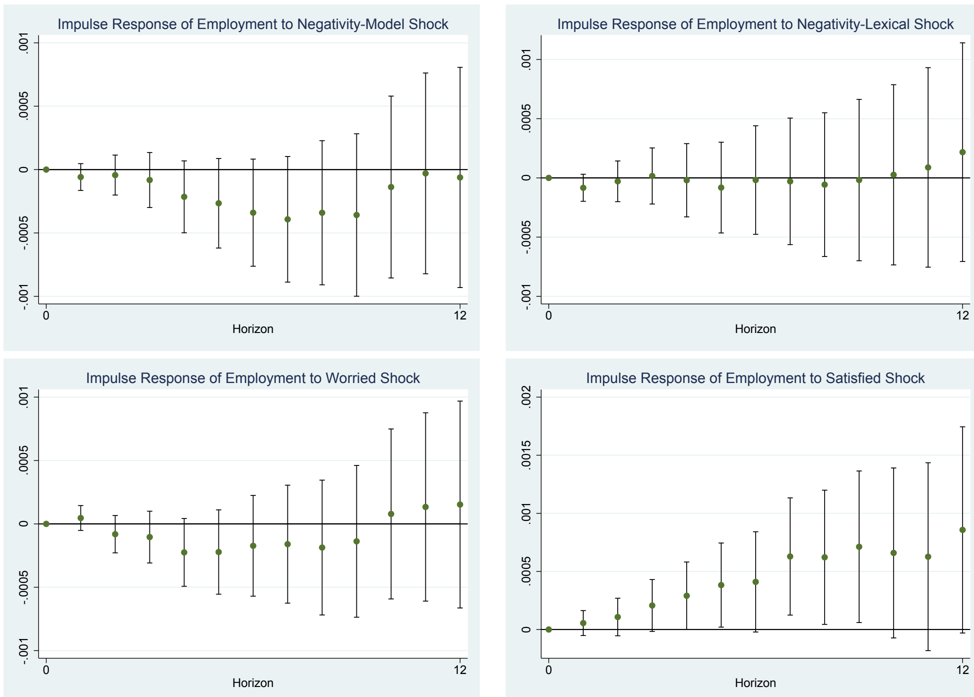
Figure 4: Heatmap of Correlations Between Sentiment Measures and Economic Outcomes



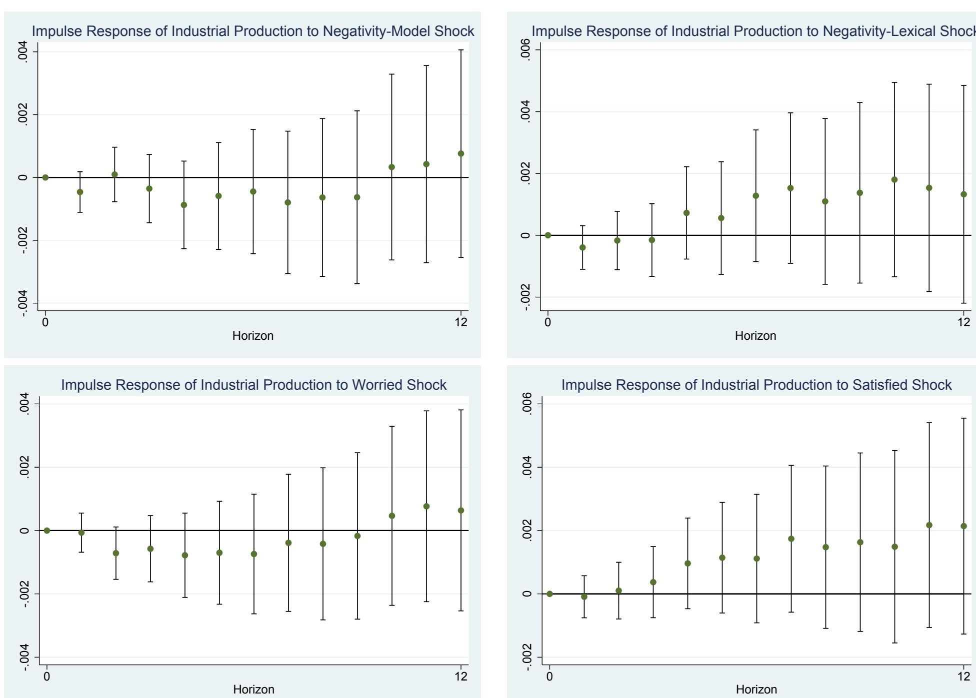
Notes: Shown are correlations between the four sentiment scores (negativity-model, negativity-lexical, worried, and satisfied) with the six economic variables (the federal funds rate, real consumption, industrial production, employment, the PCE price index, and the S&P 500 index) assessed in the study.

Figure 5: Impulse Response of News Sentiment Shock on Economic Activity

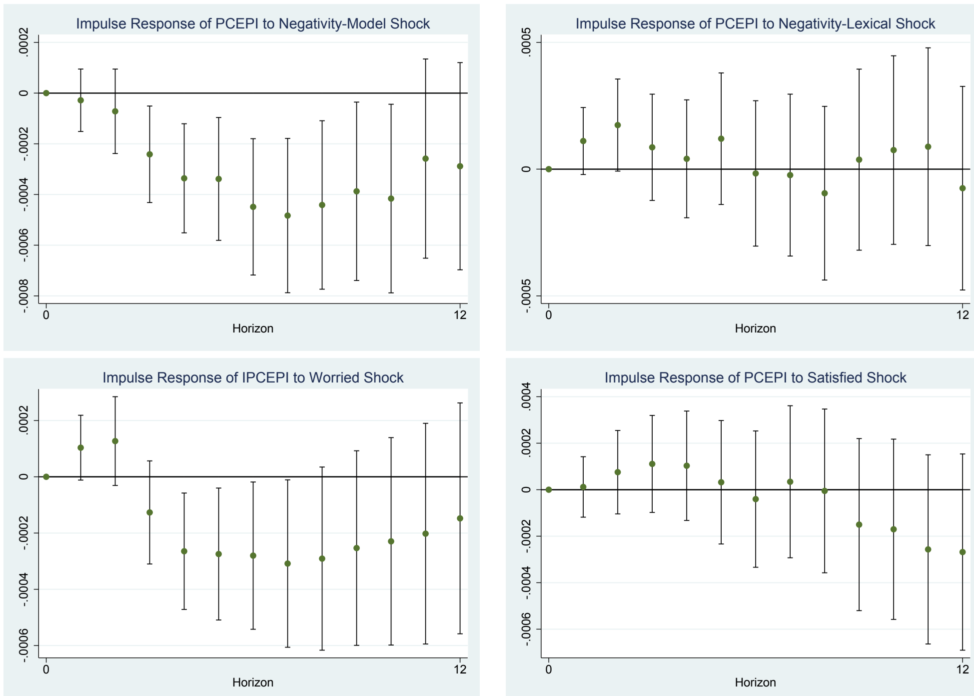
Employment



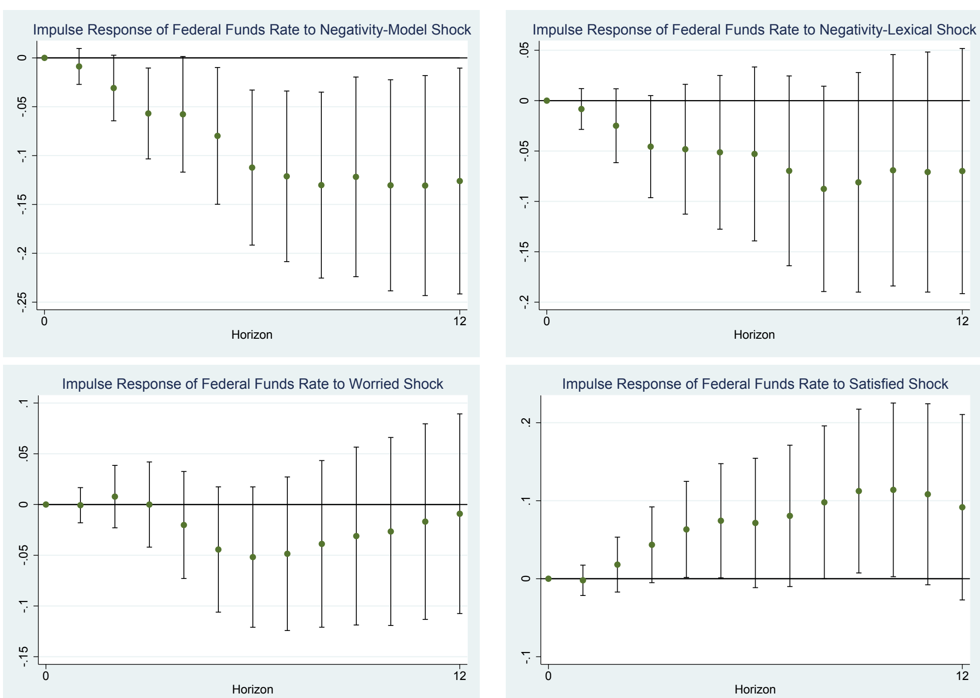
Industrial Production



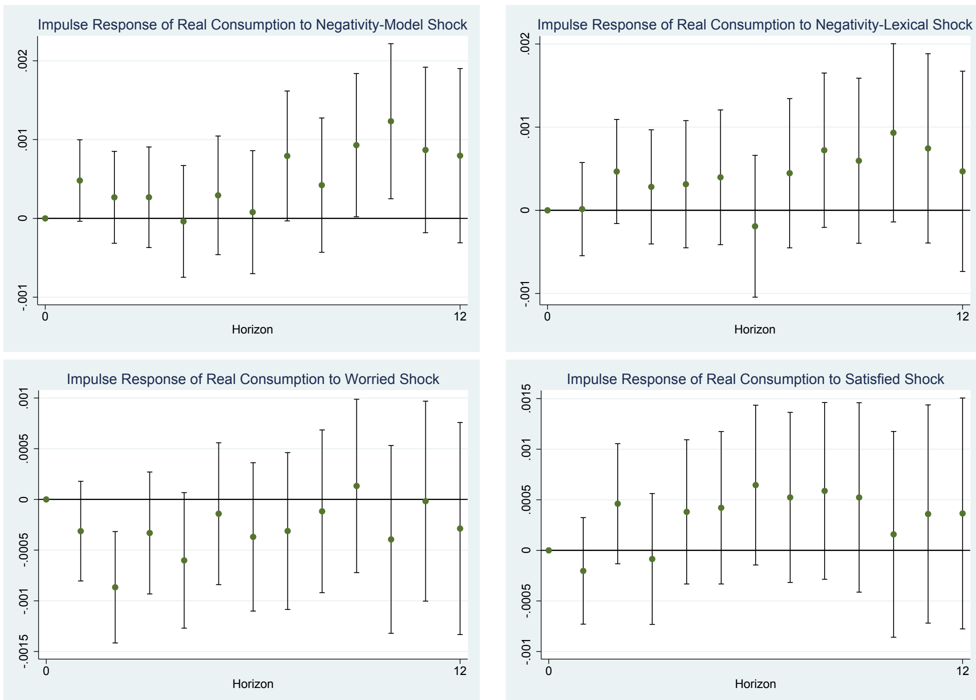
PCE Price Index



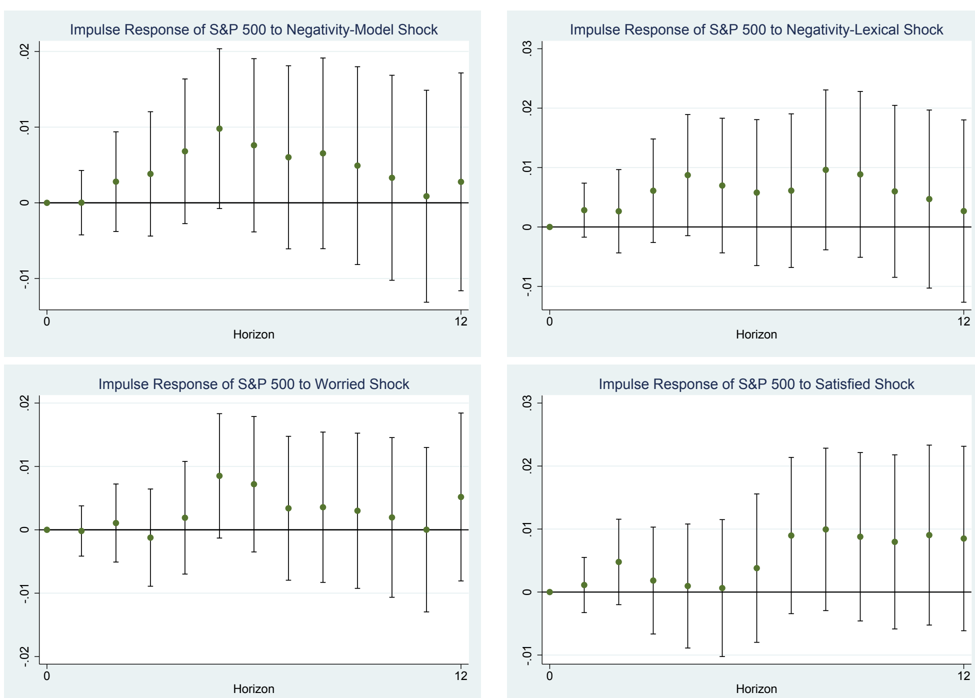
Federal Funds Rate



Real Consumption



S&P 500



Notes: Plotted are impulse responses from four different news sentiment shocks (negativity-model, negativity-lexical, worried, and satisfied) on six economic variables (the federal funds rate, real consumption, industrial production, employment, the PCE price index, and the S&P 500 index). Each news sentiment shock is constructed as the component of the news sentiment series that is orthogonal to current and 12 lags of economic activity as well as 12 lags of itself (in months). Plotted are the point estimates and 90 percent confidence bands.

Table 1

Most Common Words Associated with Selected Emotions (from Kanjoya Lexicon)

Emotion	Associated Words
negative	wrong, problem, difficult, weak, worst, disturbing, concerned, terrible, disappointing, bad
positive	steady, strong, successful, nice, excellent, glad, outstanding, tremendous, healthy, helpful
worried	worried, concerned, afraid, nervous, scared, anxious, dangerous, careful
satisfied	comfortable, impressed, satisfied, content, calm, delighted, pleased, steady, stable
confident	determined, successful, accomplish, confident, achieve, strong, forward, progress
thoughtful	question, wonder, curious, interesting, consider, explore, decision
optimistic	optimistic, potential, stronger, future, expectations, confident, growth, stronger

Table 2

Correlation Matrix

	Sentiment Negativity-Model	Negativity-Lexical	Worried	Satisfied
Negativity-Model	1.000	.	.	.
Negativity-Lexical	0.338	1.000	.	.
Worried	0.325	0.284	1.000	.
Satisfied	-0.491	-0.091	-0.084	1.000

Table 3: Predictive Accuracy: Baseline vs. Baseline w/Sentiment

	Neg(Mod)	Neg(Lex)	Satis	Worr	PC	ALL
Employment						
F(p)	0.48	0.54	0.42	0.21	0.06	0.01
AIC	10.6	5.0	9.6	5.1	5.1	-200.2
BIC	61.3	55.6	60.3	55.7	207.6	609.8
GW	-0.0013***	-0.0017***	-0.0017***	-0.0014***	-0.0030***	-0.0060***
Industrial Production						
	Neg(Mod)	Neg(Lex)	Satis	Worr	PC	ALL
F(p)	0.42	0.57	0.80	0.97	0.49	0.00
AIC	9.0	6.0	15.5	20.0	31.8	-248.5
BIC	59.6	56.6	66.2	70.6	234.3	561.6
GW	-0.0005	-0.0021	-0.0023	-0.0014	-0.0064*	-0.0177*
PCE Price Index						
	Neg(Mod)	Neg(Lex)	Satis	Worr	PC	ALL
F(p)	0.00	0.00	0.00	0.39	0.00	0.00
AIC	-37.4	-99.9	-30.3	5.7	-41.8	-368.2
BIC	13.2	-49.3	20.3	56.3	160.7	441.8
GW	-0.0007***	-0.0005	-0.0004***	-0.0003***	-0.0010**	-0.0018***
Federal Funds Rate						
	Neg(Mod)	Neg(Lex)	Satis	Worr	PC	ALL
F(p)	0.53	0.48	0.81	0.00	0.00	0.00
AIC	11.4	-1.2	15.3	-65.6	-51.4	-325.3
BIC	62.0	49.4	65.9	-15.0	151.1	484.8
GW	-0.2686***	-0.3058***	-0.3032***	-0.3027***	-0.4642***	-0.7214***
Real Consumption						
	Neg(Mod)	Neg(Lex)	Satis	Worr	PC	ALL
F(p)	0.41	0.98	0.76	0.46	0.36	0.02
AIC	8.1	20.5	14.9	7.8	19.4	-196.0
BIC	58.8	71.1	65.5	58.4	221.9	614.0
GW	-0.0019***	-0.0023**	-0.0024**	-0.0020**	-0.0039**	-0.0075**
S&P500						
	Neg(Mod)	Neg(Lex)	Satis	Worr	PC	ALL
F(p)	1.00	0.98	0.91	0.18	0.51	0.00
AIC	21.9	18.5	17.7	1.5	33.9	-300.4
BIC	72.5	69.2	68.3	52.1	236.4	509.6
GW	-0.0191**	-0.0252***	-0.0228**	-0.0241***	-0.0415***	-0.0880**

Notes: Shown are difference in statistical measures of predictive accuracy (AIC, BIC, and GW) between the baseline model and the baseline model with a corresponding sentiment measure. A negative value indicates a decline in the statistic between the baseline model and baseline model with sentiment. F(p) indicates the p-value of the F-statistic of the joint test of coefficients of all 12 lags of the sentiment equalling zero. Significance stars */**/*** indicate significance at the 10/5/1 percent level for the GW test. The baseline model is a 12-month ahead forecast of an economic variable on 12-month lags of itself and each of the five economic variables studied (i.e., federal funds rate, log real consumption, log industrial production, employment, log PCEPI, and the S&P 500 index). The baseline with sentiment model includes the baseline model plus 12 lags of a sentiment score. The first four columns include 12 lags of negativity model, negativity lexical, satisfied, and worried, respectively. The fifth column includes 12 lags each of the first three principal components of all sentiment scores available from Kanjoya. The sixth column includes 12 lags each of all sentiment scores from Kanjoya: angry, sad, upset, happy, optimistic, confused, thoughtful, confident, stressed, satisfied, worried, annoyed, appreciative, excited.

Table 4: Predictive Accuracy: Extended Baseline vs. Extended Baseline w/Sentiment

			Employment			
	Neg(Mod)	Neg(Lex)	Satis	Worr	PC	ALL
F(p)	0.07	0.02	0.72	0.15	0.06	0.00
AIC	-3.5	-18.0	13.1	1.3	-4.9	-409.7
BIC	47.1	32.6	63.8	51.9	197.7	400.3
GW	-0.0014**	-0.0015**	-0.0017***	-0.0015***	-0.0045***	-0.0045***
			Industrial Production			
	Neg(Mod)	Neg(Lex)	Satis	Worr	PC	ALL
F(p)	0.01	0.02	0.86	0.63	0.08	0.00
AIC	-13.1	-15.7	15.8	11.2	-3.4	-492.5
BIC	37.5	35.0	66.5	61.8	199.1	317.5
GW	-0.0016	-0.0017**	-0.0015**	-0.0016**	-0.0123**	-0.0123**
			PCE Price Index			
	Neg(Mod)	Neg(Lex)	Satis	Worr	PC	ALL
F(p)	0.01	0.00	0.00	0.55	0.00	0.00
AIC	-16.6	-104.7	-33.0	6.6	-38.9	-480.7
BIC	34.0	-54.1	17.6	57.2	163.6	329.4
GW	-0.0004***	-0.0004***	-0.0004***	-0.0004***	-0.0012***	-0.0012***
			Federal Funds Rate			
	Neg(Mod)	Neg(Lex)	Satis	Worr	PC	ALL
F(p)	0.99	0.70	0.99	0.00	0.00	0.00
AIC	20.5	3.5	21.2	-63.1	-53.2	-466.1
BIC	71.2	54.1	71.8	-12.5	149.3	344.0
GW	-0.2836**	-0.3101**	-0.3213**	-0.3253**	-0.6284***	-0.6284***
			Real Consumption			
	Neg(Mod)	Neg(Lex)	Satis	Worr	PC	ALL
F(p)	0.01	0.66	0.69	0.50	0.03	0.00
AIC	-12.3	12.2	12.6	6.7	-18.7	-388.9
BIC	38.3	62.8	63.2	57.3	183.8	421.1
GW	-0.0026**	-0.0026**	-0.0022**	-0.0023**	-0.0061***	-0.0061***
			S&P500			
	Neg(Mod)	Neg(Lex)	Satis	Worr	PC	ALL
F(p)	0.93	0.91	0.94	0.33	0.11	0.00
AIC	17.6	15.9	18.1	4.3	4.6	-472.0
BIC	68.2	66.5	68.8	54.9	207.1	338.1
GW	-0.0260***	-0.0287***	-0.0232**	-0.0288***	-0.0754***	-0.0754***

Notes: Shown are difference in statistical measures of predictive accuracy (AIC, BIC, and GW) between the extended baseline model and the extended baseline model with a corresponding sentiment measure. A negative value indicates a decline in the statistic between the baseline model and baseline model with sentiment. F(p) indicates the p-value of the F-statistic of the joint test of coefficients of all 12 lags of the sentiment equalling zero. Stars */**/*** indicate significance at the 10/5/1 percent level for the GW test. The extended baseline model is a 12-month ahead forecast of an economic variable on 12-month lags of itself and each of the five economic variables studied (i.e., federal funds rate, log real consumption, log industrial production, employment, log PCEPI, and the S&P 500 index) plus 12 lags of the Conference Board's consumer confidence index and the Michigan consumer sentiment index. The extended baseline with sentiment model includes the extended baseline model plus 12 lags of a sentiment score. The first four columns include 12 lags of negativity model, negativity lexical, satisfied, and worried, respectively. Column 5 includes 12 lags each of the first three principal components of all sentiment scores available from Kanjoya. Column 6 includes 12 lags each of all sentiment scores from Kanjoya: angry, sad, upset, happy, optimistic, confused, thoughtful, confident, stressed, satisfied, worried, annoyed, appreciative, excited.

Table 5: Predictive Accuracy: Extended Baseline vs. Baseline w/Sentiment

	Neg(Mod)	Neg(Lex)	Satis	Worr	PC	ALL
Employment						
AIC	0.1	-5.6	-0.9	-5.5	-5.4	-210.8
BIC	-50.5	-56.2	-51.6	-56.1	95.8	498.0
GW	0.0003	-0.0002	-0.0002	0.0001	-0.0014***	-0.0045***
Industrial Production						
	Neg(Mod)	Neg(Lex)	Satis	Worr	PC	ALL
AIC	-9.3	-12.3	-2.8	1.7	13.5	-266.8
BIC	-59.9	-62.9	-53.4	-48.9	114.7	442.0
GW	0.0049	0.0033	0.0031*	0.0040	-0.0011	-0.0123**
PCE Price Index						
	Neg(Mod)	Neg(Lex)	Satis	Worr	PC	ALL
AIC	-40.4	-102.9	-33.3	2.7	-44.8	-371.2
BIC	-91.0	-153.5	-83.9	-47.9	56.4	337.6
GW	-0.0001	0.0001	0.0002	0.0003	-0.0004***	-0.0012***
Federal Funds Rate						
	Neg(Mod)	Neg(Lex)	Satis	Worr	PC	ALL
AIC	-3.9	-16.4	0.0	-80.8	-66.6	-340.5
BIC	-54.5	-67.1	-50.6	-131.5	34.6	368.3
GW	-0.1756	-0.2128*	-0.2102*	-0.2096*	-0.3712***	-0.6284***
Real Consumption						
	Neg(Mod)	Neg(Lex)	Satis	Worr	PC	ALL
AIC	-19.2	-6.8	-12.5	-19.5	-7.9	-223.4
BIC	-69.8	-57.5	-63.1	-70.2	93.3	485.4
GW	-0.0005	-0.0009**	-0.0010**	-0.0006	-0.0024***	-0.0061***
S&P500						
	Neg(Mod)	Neg(Lex)	Satis	Worr	PC	ALL
AIC	2.2	-1.1	-2.0	-18.2	14.2	-320.1
BIC	-48.4	-51.7	-52.6	-68.8	115.5	388.7
GW	-0.0065	-0.0126**	-0.0102*	-0.0115**	-0.0289***	-0.0754***

Notes: Shown are difference in statistical measures of predictive accuracy (AIC, BIC, and GW) between the extended baseline model and the baseline model with a corresponding sentiment measure. A negative value indicates a decline in the statistic between the baseline model and baseline model with sentiment. Stars */**/*** indicate significance at the 10/5/1 percent level for the GW test. The extended baseline model is a 12-month ahead forecast of an economic variable on 12-month lags of itself and each of the five economic variables studied (i.e., federal funds rate, log real consumption, log industrial production, employment, log PCEPI, and the S&P 500 index) plus 12 lags of the Conference Board's consumer confidence index and the Michigan consumer sentiment index. The baseline with sentiment model includes the baseline model plus 12 lags of a sentiment score. The first four columns include 12 lags of negativity model, negativity lexical, satisfied, and worried, respectively. The fifth column includes 12 lags each of the first three principal components of all sentiment scores available from Kanjoya. The sixth column includes 12 lags each of all sentiment scores from Kanjoya: angry, sad, upset, happy, optimistic, confused, thoughtful, confident, stressed, satisfied, worried, annoyed, appreciative, excited.

A A. Examples of Scored Newspaper Articles

A.1 Article 1

"Analysis: Fewer to get health insurance" (*Chicago Tribune*, July 25, 2012)

Negativity-Model: 0.995 (99th percentile)

Negativity-Lexical: -0.016 (28th percentile)

Worried: 0.145 (35th percentile)

Satisfied: 0.024 (2nd percentile)

Fewer Americans will likely get health insurance over the next decade under President Barack Obama's health care law as a result of the Supreme Court's decision to limit it, according to a new analysis of the landmark ruling.

And the court's decision to allow states to opt out of a major expansion of the government Medicaid insurance program for the poor could also save taxpayers $84 billion by 2022, the nonpartisan Congressional Budget Office estimates.

The new projections confirm that the court's ruling will not fundamentally alter the law Obama signed in 2010.

The budget office, which Congress relies on to study the impact of proposed legislation, estimates the Affordable Care Act will still reduce the number of Americans without health coverage by some 30 million over the next decade.

And the law will continue to lower the federal deficit because costs of expanding coverage are offset by new taxes and cuts in other federal spending. Repealing the law, in contrast, would increase the deficit by $109 billion, budget analysts said.

But the Congressional Budget Office estimates that the court's decision will lead to major variations across the country in how states implement the law.

The law originally required states to expand their Medicaid programs in 2014 to cover all Americans making less than 138 percent of the federal poverty level, or about $15,400 for a single adult. Today, Medicaid is mostly limited to poor families, elderly and the disabled.

A.2 Article 2

"U.S. income growth robust in 1998" (*Times-Picayune*, April 28, 1999)

Negativity-Model: 0.221 (3rd percentile)

Negativity-Lexical: 0.030 (99th percentile)

Worried: 0.173 (49th percentile)

Satisfied: 0.113 (19th percentile)

Americans' incomes climbed 4.4 percent in 1998, significantly outpacing inflation. In every state, per capita incomes grew faster than prices.

Nationally, average income for America's 270.3 million people last year was $26,412, up 4.4 percent from 1997, the Commerce Department said Tuesday. The figures include not only wages and salaries, but other sources of income such as investment earnings and government benefits.

That per capita income growth was actually a bit slower than the 4.7 percent increase in 1997.

But the inflation rate – as measured by the government's price index for personal consumption expenditures – declined from 2 percent in 1997 to 0.8 percent last year. Factoring that in, Americans' buying power – per capita income growth after inflation – jumped 3.6 percent in 1998 compared with 2.7 percent in 1997.

By state, per capita incomes in 1998 ranged from $37,598 in Connecticut to $18,958 in Mississippi. Growth rates in per capita income ranged from 7.8 percent in North Dakota, where farmers hurt by falling international commodity prices got relief payments from the government, to 2.1 percent in Hawaii, where tourism has fallen off because of economic troubles in Asia.

After North Dakota, the states with the fastest per capita income growth rates were Colorado, 6.1 percent; Washington, 5.7 percent; Texas, 5.3 percent, and Vermont and Massachusetts, 5.0 percent.

Besides Hawaii, other states with relatively slow per capita income growth last year included Wyoming, 2.5 percent; Nevada and Montana, 2.6 percent, and Alaska, 2.8 percent. Nevada has experienced strong economic growth in recent years, but the effects on its per capita income have been diluted by simultaneous rapid population growth.

Increased earnings from service-sector jobs – especially in finance and real estate – was the driving force behind high per capita income growth in many states, such as Massachusetts and Colorado.

Other states such as Texas and Washington have seen growth in relatively high-paying technical jobs, such as in the computer and telecommunications industries.

Nationally, unemployment is near a 29-year low, and brisk spending by Americans has been largely responsible for keeping the U.S. economy growing while other countries have been hurt by a global economic downturn that started two years ago in Asia.

States heavily dependent on farming, such as North Dakota, have been among those feeling the worst effects of the international crisis that has pulled down U.S. exports.

But some farming states' per capita income growth was pushed to among the best in the nation by billions of dollars in relief payments Congress and President Clinton approved late last year.