Economic Conditions, Economic Perceptions, and Media Coverage of the United States Economy

This paper consists of two distinct sections. First, we examine a particular coding scheme for measuring the tone of media coverage of the economy; and compare the utility of training data sets produced by different sets of coders. We utilize: 1) undergraduates at Penn State University; 2) coders on CrowdFlower drawn from across the planet; and 3) coders on CrowdFlower restricted to those from the United States. We use datasets produced by these three sets of coders to train a classifier on coding articles from the *New York Times* from 1948 to 2014. In the second section of the paper we classify articles from the *New York Times* using an alternate training data set. And we examine two aspects of media coverage of the economy. First, we look at what objective economic indicators drive the content of media coverage of the economy. Second, we look at the impact of media coverage of the economy on economic perceptions. We pay special attention to whether media coverage is driven by changes in different measures of the state of the economy: including unemployment, inflation, personal income, and the stock market. We then examine the impact of this media coverage of the economy on economic perceptions using the index of Consumer Sentiment, as well as other available survey data. Our analysis covers media coverage and public perceptions of the economy in the United States over a fifty year period, and economic perceptions over a thirty year period.

October 8, 2014

Pablo Barberá
New York University

Amber Boydstun
University of California, Davis

Suzanna Linn
Penn State University

Jonathan Nagler
New York University

Prepared for delivery at the 5th Annual Text as Data Conference, October 10-11, 2014. A very similar version was delivered at the 2014 EPSA and APSA conferences. The authors can be reached at pablo.barbera@nyu.edu, aboydstun@ucdavis.edu, slinn@la.psu.edu and jonathan.nagler@nyu.edu, respectively.

Our first goal in this paper is to determine which aspects of the macro-economy determine the tone of newspaper coverage of the economy. Our second goal is to determine the distinct impact of different aspects of the macro-economy, and the tone of media coverage of the economy, on individuals' perceptions of the state of the economy. We do this by coding all newspaper articles that might be about the state of the United States economy appearing in the New York Times from 1948 through 2010. We create a monthly series of media-tone, and then model this series as a function of assorted measures of the macro-economy (unemployment, inflation, personal income, and the stock market). For a subset of this period where we have the index of consumer sentiment available, we then examine the impact of media tone and the economy on consumer sentiment.

1 Motivation

Economic inequality has risen dramatically in the United States over the last 40 years. Concern that the political system has failed to represent the interests of poorer Americans has risen in tandem. Poorer Americans participate less, and often seem to discount their own interests when they do participate. Elected officials thus have little incentive to respond to their needs.

As the story of electoral reward and punishment usually goes, voters look to the national economy for evidence as to whether the incumbent president (or party) is managing the economy in their interest and reward or punish the incumbent in accord with this information (Key 1966, Hibbs 2012). But economic inequality presents a challenge to voters. Real mean family income has increased by 30.2% since 1980. Yet the average voter sitting in the bottom income quintile would have experienced real income growth of *negative* 7.3% (absent life-cycle effects), while his or her counterpart in the very top

5% of the income distribution saw real growth of over 93% over that period.[1] If voters judge incumbents based on the performance of the aggregate economy, those voters at the bottom of the income distribution have ceded their role as "rational god(s) of vengeance and reward" (Key 1964), at least so far as that vengeance or reward is based on the self-interest of the voter.

Yet in recent research two of us find that voters in the bottom 40% of the income distribution pay relatively little attention to income growth of their own income quintile, and more attention to aggregate income growth (Linn & Nagler 2014). Further, this research suggests that over the most recent 10 presidential elections, those occurring in this era of rising economic inequality, economics appears to motivate voter behavior less than in the past. In an even more troubling finding, Larry Bartels (2008) finds that the voting behavior of Americans at all positions in the income distribution reflects the economic experiences of the wealthiest 5% of Americans.

We conjecture that these behavioral findings suggesting that lower income voters have been ignoring their own economic misfortune when voting are a result of the nature of media coverage of the economy and its influence on economic evaluations that voters make. Briefly, we suspect that the tone of media coverage is biased towards the economic performance of those at the top of the income distribution, and that it provides relatively little information about the economic performance of those at the bottom of the income distribution. This information then colors individuals' perceptions of the economy, and can move the economic evaluations of lower income voters away from a more accurate view of the economic performance of their group. These economic evaluations in turn influence voter choices. In this way, the media truly mediate the impact of the economic experience of voters in different groups by acting on evaluations of the economy.

[1] Values computed from Census Bureau Family Income data by authors.

Media provide voters with information on and evaluation of the national economy, reporting either positive or negative information or opinions on the state of the economy. And media *can* also provide information on the circumstances facing voters in different economic groups, or media can provide information that simply informs the voter of variations in economic experiences. For instance, the media could report on aggregate income growth over the last year, income growth in the bottom quintile, or rising income inequality.

A large body of research has shown that the tone of media coverage of the economy influences economic evaluations by the mass public (Ansolabehere, Meredith & Snowberg 21012, Blood & Phillips 1995, De Boef & Kellstedt 2004).[2] Thus one potential explanation for the perverse behavioral findings that motivate our research is that tone reflects only the performance of the aggregate economy, or the performance of those in the top 5% of the income distribution, causing perceptions of economic conditions – even the economic conditions of one's group – to be disproportionately influenced by aggregate economic performance, or economic performance of the top 5%. A related explanation is that media coverage exhibits relatively little breadth, i.e., that scalar measures of the national economy so dominate media coverage of the economy that media coverage can be reduced to nothing more than a measure of the national economy, and that it contains no information about the variance in economic performance across specific groups. In both cases voters would be left incapable of punishing incumbents for poor economic performance for their own income group and thus of demanding accountability from government policy makers.

In this paper we do not examine measures of the economy related specifically to economic inequality, but we first test our ability to code media tone, and demonstrate that we can do this successfully and show that media tone is a function distinct real economic

[2]See also Hester (2003), Hetherington (1996), Goidel (1995), Mutz (1992, 1994), Pruitt (1989), Sanders (1993), Stevenson (1994), and Tims (1989) for additional research showing that media coverage of the economy influences economic evaluations by the mass public.

indicators.

2 Data

Our analysis of media coverage requires developing new methods to measure how news outlets decide to inform citizens about the state of the economy. In particular, we are interested in measuring the "tone" or "sentiment" of media coverage, which we define as the extent to which a story suggests that the economy is performing well. Previous studies have implemented different strategies to capture this variable. One approach is to develop dictionaries of positive or negative words, and then count their appearance in newspaper articles (Young & Soroka 2012, De Boef & Kellstedt 2004). However, these methods generally have low accuracy, as we show below. A different strategy is to manually code the tone of all stories (or a random sample), but this becomes impractical as the period of analysis increases. To overcome these two limitations, we rely on recent developments in the fields of statistics and computer science (Hastie et al. 2009) to predict the tone of our entire corpus by training a machine learning classifier on a random sample of articles.

2.1 Data Collection

We began by collecting a sample of stories about the U.S. economy from 1948 to 2010 that appeared in the first 20 pages of *The New York Times* .[3] We retrieved a population of potentially relevant stories using the Proquest Archive of Historical Newspapers. Our query included terms related to economic indicators and the general state of the economy, such as "unemployment", "inflation", "GDP", "stock market", "gas price", etc., and was aimed

[3] *The New York Times* has not been consistently sectioned. We draw from the first 20 pages of the paper in order to approximate the first section of the newspaper, that with the widest readership.

at identifying stories relevant to all aspects of the domestic economy while minimizing international or other irrelevant stories.[4] Our search returned 73,155 stories, of which 9,336 appeared in the first 20 pages. From these, we excluded 1,364 stories that mentioned other countries in their headline (without mentioning "U.S." or "United States") in order to further restrict our analysis to articles that focus on the U.S. economy.[5] Our final sample size is 8,072 stories.

Figure 1 shows that *The New York Times* published a consistent and fairly large number of stories meeting our retrieval conditions for (perhaps) being about the U.S. economy in the first 20 pages over this period, with an average of 126.1 stories per year and 10.6 per month (standard deviation of 8.3). With the exception of the two months the paper was on strike (September and October of 1978), the minimum number of stories on the economy per month was 1, and the maximum was 76.

After collecting the stories from ProQuest in PDF format, we used optical character recognition software (OCR) to convert them into machine-readable text.[6] The final step in our data collection was to identify sentence breaks using regular expressions, and use these breaks to extract the first five sentences from each article.

[4]The exact query was: ab(unemployment OR inflation OR "consumer price index" OR GDP OR "gross domestic product" OR "interest rates" OR "household income" OR "per capita income" OR "stock market" OR "federal reserve" OR "consumer sentiment" OR recession OR "economic crisis" OR "economic recovery" OR globalization OR outsourcing OR "trade deficit" OR "consumer spending" OR "full employment" OR "average wage" OR "federal deficit" OR "budget deficit" OR "gas price" OR "price of gas" OR "deflation" OR "existing home sales" OR "new home sales" OR "productivity" OR "retail trade figures" OR "wholesale prices") AND "United States".

[5]We used the list of country names in standard format available in the "countrycode" package for R (2014). This unfortunately does not remove articles using any variant of a foreign country name. For instance, articles about the "British" pound, would still be included.

[6]The OCR software we used was Abby FineReader. A visual analysis of a random sample of articles showed that this tool was able to preserve most of the text of the articles as it was published, with only minor formatting differences.

2.2 Creating a Training Dataset

Before proceeding, we discuss alternative concepts of "tone." We could proceed with two fundamentally different models. One model would treat tone as an objective truth: that there is a unique value of tone associated with a new story, and that value is fixed for all readers of the story. The competing model treats tone as something that is reader-specific; allowing tone to be a characteristic inferred by the reader, and allowing for heterogeneity across readers in how they perceive tone. In the creation of training datasets that we describe below we explicitly choose the latter approach. Coders are given no guidance as to what facts would constitute bad news or good news about economic performance. The coders are explicilty told "your job is to judge whether the sentence gives YOU an indication about how the economy is performing." This suggests that the choice of coders could matter greatly, and that we should be concerned with the representativeness of our coders for the population we wish to study.

We explored different alternatives to create a dataset that can be used to train a machine learning classifier for our variables of interest. First, we trained six undergraduate students and one graduate student to manually code the approximately 5 lead paragraphs of 400 articles.[7] This same set of articles, as well as 400 additional stories, were also labeled by crowd coders on Crowdflower (2014). Finally, we used the location filtering options in Crowdflower to have coders located only in the U.S. also label these 400 articles. The comparison across these different sources of coding allow us to explore some of the trade-offs associated to the construction of a manually annotated dataset.

The coding task was identical in all three cases. Multiple coders (at least 3) coded each article. All coders were asked to answer a series of questions with regard to each

[7] The data we use in this draft corresponds to the 282 sentences coded up to October 7th, 2014.

sentence (in some cases sentences). They were first asked: "Does the sentence provide some indication about how the U.S. economy is performing?" If they answered "no" or" not sure", they were presented with the next sentence(s) in the article. If they answered "yes" to the first question, they were presented with three additional questions. The second question asked them whether the sentence indicated the performance of the economy was positive, negative, neutral, mixed, or whether they were not sure. The final two questions asked coders to attribute their response to the facts in the story and/or the language used by the journalist: "Was your answer to Q2 based in part (or entirely) on the objective facts presented in the sentence (i.e., statistics, or descriptions or quotes that clearly reveal levels or changes of economic indicators)?" and "Was your answer to Q2 based in part (or entirely) on how the journalist chose to present the facts (i.e., the language the journalist used, or whether they chose to present the fact in a positive or negative manner)?" Yes, no, and not sure were the permitted response options. Our goal in asking the last two questions was to determine the source of the tone evaluation and responses are used in conduction with revealed tone to test a set of hypotheses including whether the tone based on objective facts only is more responsive to economic indicators than the tone based only on the language the journalist used and whether consumer sentiment responds more to tone based on language only more strongly than to tone based only on factual information. In this draft of the paper we focus on the first two variables and leave the discussion of the other two for future versions.

Table 1 reports a summary of the levels of intercoder reliability within each source of codings and across sources. First, the diagonal values indicate the average pairwise agreement rate within each source of coding, that is, the expected proportion of respondent pairs who gave the same answer to each question. We find that undergraduates give more consistent answers than crowd coders, and that in all three cases we get relatively high levels of agreement *within each group*. Note that the percent agreement if coders were giving random answers would be 0.33 in the relevance question and 0.25 in the tone question.

Second, the values off the diagonals indicate the percent agreement on the modal response to each question across different sources of codings. We find that, despite the high levels of within-source agreement, responses from the crowd are less consistent with responses from undergraduate students that we expected. The rates of agreement are even lower when we focus on the U.S. crowd coders.

This table thus suggests that responses given by undergraduate students are of better quality than those obtained by aggregating responses from 3 crowd coders.[8] At the same time, however, relying on crowd coders is significantly cheaper and faster, which presents researchers with a trade-off. It also shows that filtering by location might not necessarily be a good idea. We plan to explore this result further in the future to understand whether it is because of the demographic characteristics of who is coding in the U.S., the timing of the coding, or the number of different coders (which was lower in the U.S., perhaps due to the smaller sample of active coders in this country).

[Table 1 Here]

2.3 Estimating Sentiment Using Machine Learning

To estimate a model of relevance and tone that could be applied to a larger dataset we first pre-processed the text of the training data by removing English stopwords, words with

[8] We note that given the nature of the coding scheme, that coders are being asked their own interpretation of the articles - there are multiple interpretations of this table. The undergraduate coders may not be of higher 'quality' than other coders - there may simply be less variation between undergraduates in how they interpret economic news.

less than two and more than 20 characters, and words that appear in more than 80% of the stories.[9] For the tone classifier, we only kept stories in which some sentences were coded as having positive or negative tone. Then, we experimented with different machine learning classifiers in the `scikit-learn` library for python (Pedregosa et al. 2011), such as SVM, ElasticNet, Naïve Bayes, and regularized logistic regression; and varying the number of features and n-grams.[10] Our measure of model fit was cross-validated accuracy.[11] We found that a regularized regression with L2 penalty (ridge regression) using up to trigrams and keeping the top 50,000 most frequent n-grams as features maximizes accuracy, and thus all the results we report were estimated using this type of classifier.

Table 2 displays the performance of this classifier trained with different data sources. We find that, despite the smaller sample of sentences coded by undergraduate students, a classifier trained with this data appears to outperform the others: its accuracy for the relevance question is 0.80 and for the tone question is 0.73.[12] However, note that when making comparisons we also need to take into account what the modal category is. This allows us to examine to what extent the prior (the observed proportion in the data) is dominating the classifier, or if the classifier is really finding some signal. Based on this comparison, we see that the tone classifier trained with data coded by undergraduates performs only slightly better than the prior. The classifier trained with crowd-coded data,

[9] This pre-processing may be too severe for our purposes. As we are training based on bi-grams and tri-grams, excluding common words (uni-grams) could be overly restrictive: by excluding a word such as 'economic' we would also be excluding 'poor economic'.

[10] In future versions of the paper we plan to explore classifiers that allow us to incorporate priors over sentence tone, recognizing that sentences in a given article are not independent. Rather we could consider this to be a multi-level data problem: where sentences are nested within an article; and the article has some overall level of tone and sentences are drawn from a distribution centered on that overall level.

[11] We use five-fold cross-validation: we split the data in five random samples, train the classifier on four of them, predict the labels for the remaining 20%, and repeat for each fold. The standard errors we report in the text are based on the variability of the accuracy (% sentences with correct prediction divided by the total number of sentences in each fold). The confusion matrix aggregates the predictions and observed values over all five folds.

[12] Note that accuracy here is a measure of how well we match the modal category chosen by the set of coders. So we are losing a measure of what could be inherent noise in the data.

on the contrary, performs four percentage points better than the prior. This increases to a 9-point improvement when we subset the crowd-coded data by confidence in the response, and keep only those above the median. Finally, perhaps not surprisingly given what we showed in Table 1, the classifier trained with data coded by U.S. coders has extremely poor performance.

[Table 2 Here]

2.4 Switch to Alternate Training Data

Given the poor performance of these different classifiers, the results we report after this point use a classifier trained in a different dataset – a sample of 1,900 stories from *The New York Times*, *Wall Street Journal* (abstracts only), *The Washington Post*, and *USA Today*, published between 1980 and 2011. For this sample, we trained student coders to categorize each story according to whether the economic news was *primarily* positive or negative (or neutral). Each story was coded by a single coder and then checked by a more experienced "master" coder.[13] Note that this is thus a training dataset: 1) trained only by student coders; 2) trained on a time period that does not match the period we wish to classify data from; and 3) is based on coding of a larger section of text from the article rather than coding of individual sentences; and 4) is based on a wider set of media sources.

The accuracy of a classifier trained with this data was 73% (+/- 2%), as shown in the confusion matrix in Table 3. The classifier therefore performs better than random (50% accuracy) and the modal category (60% of the articles in the training set have positive tone). It also outperforms existing dictionary approaches such as LexiCoder (Young &

[13]Intercoder reliability was high, with 93.5% agreement and both Cohen's Kappa and Krippendorff's Alpha scores of 0.896 based on a sample of 200 stories integrated into coders' files without their knowledge.

Soroka 2012), with 63.2% accuracy in our labeled data. Furthermore, we find that the features with the highest and lowest estimated coefficients in the ridge regression correspond to our expectations regarding the type of words that should appear in news stories with negative and positive tone, as we show in Table 4

[Tables 3 and 4 Here]

2.5 Sentiment Data

We measured the tone of media coverage by applying our classifier to the full sample of articles about the U.S. economy from our original search. To repeat the caution above: we are now using the classifier trained on the set of approximately 1900 stories coded coded from a mix of papers from 1980 thru 2011. And it is the estimate of Tone we compute here that is used in alter sections of the paper. Specifically, we used our classifier to compute the predicted probability that each sentence in an article has a positive tone about the economy and then averaged across the sentence scores to produce an overall predicted probability for the article. Since the finest level of aggregation for most economic indicators is the month, we aggregated these probabilities by month, weighting them by the number of words in each article. Our measure of sentiment for month t is thus

$$s_t = \frac{1}{\sum_t w_{it}} \sum_i s_{it} \times w_{it} \tag{1}$$

where s_{it} is the predicted probability that article i in month t is positive and w_{it} is the total number of words in that same article. Note that aggregating probabilities instead of predicted tone allows us to propagate the uncertainty about the individual predictions to the monthly estimate, following the intuition in Hopkins and King (2010). We give more weight to longer articles for two reasons: first, article length is also an editorial decision with important implications for media coverage; second, we find that the machine learning classifier performs better the more text it has to generate a prediction.

Figure 2 displays our estimated sentiment data. The average probability a story is positive is 43% (standard deviation 5.6) from 1947–2009 with a range of 28% to 61% ($T = 756$). The *tone* of the stories tends to be slightly negative, but close to neutral. The mid 1960s, the late 1980s, and the late 1990s have the most positive average probability a given story is positive. The early 1970s, 1980s, and the most recent recessionary period show the lowest average probability a story is positive. These figures changed slightly in the more recent era, which we demarcate as 1978 forward ($T = 384$), when the data on economic perceptions is first available (see Figure 3). The average probability a story is positive is 44% with a standard deviation of 6%.

2.6 Economic Data

Three measures of economic performance are available to us and have been widely reported (monthly) over the full period covered by our economic news stories: unemployment, inflation, and stock market performance. They have the added virtue of being widely recognized by economists and generally recognized by consumers as indicators of the health of the national economy. Measures of personal disposable (and personal) income, available monthly since 1959, cover a large span of the data as well, and capture an important aspect of consumers' ability to spend.[14] We thus concentrate our attention on the ability of these

[14] Data on the S&P is from Shiller `http://www.econ.yale.edu/~shiller/data.htm`, and is inflation adjusted. When using it to model media tone it is measured in percent change of the index from the previous period. The remanning economic data was obtained through FRED. Disposable income growth is measured as (annualized) percent change in billions of 2009 chained dollars. Unemployment is seasonally unadjusted. Monthly inflation data is (annualized) percent change in the consumer price index for all goods (all urban consumers).

measures of economic performance to explain the tone of news coverage of the economy, and to help us discern which aspects of economic performance determine tone.

Our analysis of economic evaluations includes these same measures of economic performance and also the Conference Board coincident indicator index, which is composed of payroll employment in nonagricultural businesses, personal income (less transfer payments, inflation adjusted), industrial production and real manufacturing and trade sales. Our purpose here is to draw on the expertise of economists following a variety of economic indicators that have historically tracked and forecast economic performance.[15] We use this indicator to help control broadly for economic performance in order to assess the independent effect of the tone of media coverage of the economy on citizens' economic perceptions. We do not use the Conference Board index to explain tone because in doing so we would lose our ability to identify the the specific source(s) of economic performance driving media tone.

Monthly data on economic perceptions is from The University of Michigan Survey of Consumer Attitudes and is available beginning in 1978. The Survey asks respondents a set of questions related to the performance of the economy. Five questions are combined to create the Index of Consumer Sentiment, which is widely reported by the media. These include forward and backward-looking questions about personal finances, two forward-looking questions about the national economy (12 months and five years ahead), and one question asking respondents to evaluate whether it is a good time to buy major household appliances.[16] We examine the responsiveness of the ICS to both objective economic

[15] See `https://www.conference-board.org/data/bci/index.cfm?id=2160` for details on the creation and composition of the index.

[16] The five questions that comprise the index are: 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

performance and media coverage of the economy below. We also separately model evaluations of the national economy, both the 12 month and five year ahead assessments, and retrospective evaluations of the national economy.[17]

3 Relationship Among Media Tone, Real Economy, and Economic Perceptions

We report correlations between our measure of the tone of media coverage, the article count each month, our economic performance measures, and two measures of consumer sentiment in Table 5. The top panel of the table gives the correlations for 1948 thru 2010, the bottom panel gives the correlations for the more recent period 1970 thru 2010.

The two media coverage time series are negatively correlated—higher levels of coverage are associated with a lower average probability a story is positive, consistent with a media focus on covering "bad news". (This correlation is higher since 1978.) The tone of news coverage of the economy is correlated with three of our economic indicators in ways we expect: tone is more positive when the stock market is rising (ρ =0.09) and both unemployment (ρ =-0.19) and inflation (ρ =-0.14) are relatively lower. These correlations are stronger for the relationship between tone and unemployment and inflation since 1978, increasing to -0.49, and -0.20, respectively. But tone is remarkably unrelated to growth in

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, 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?" Each question is scored by calculating the percent giving favorable replies minus the percent giving unfavorable replies and adding 100. The ICS is formed by adding these scores and then dividing by a normalizing constant and adding an additional constant to correct for over time change in the sampling design.

[17] The specific survey question asks respondents: "Would you say that at the present time business conditions are better or worse than they were a year ago?"

disposable personal income (ρ =0.01) or the percentage change in the Conference Board economic indicator indices from 1959-2010 or 1978–2010. The correlation between changes in our economic indicators and the average probability a story is positive (reported for 1978–2010 only) are notable for their uniform proximity to zero. Only change in inflation has even a weak correlation with media tone (ρ =-0.20).

[Table 5 Here]

The correlations involving our two measures of evaluations of the economy (retrospective and prospective evaluations, available in the second panel of Table 5 for 1978 thru 2010) are positively related to media tone. In particular, prospective evaluations of the national economy 12 months from now are more strongly related to the average probability a story is positive ($\rho = 0.30$) than are retrospective evaluations of the economy 12 months in the past ($\rho = 0.21$). Evaluations of both sorts tend to be more negative when the number of articles is higher (ρ =-0.29 and -0.23, respectively). The correlation of evaluations with unemployment are higher than with tone (ρ=-0.34 and -0.37, for retrospective and prospective evaluations, respectively) and with the coincident index (ρ=0.56 and 0.49, respectively).

4 The Economic Roots of Media Tone

What economic indicators explain the tone of media coverage since WW II? In Table 6 we report block F-Test results from a regression of media tone on: inflation; changes in unemployment; percent change in the S&P Index; and (annualized) real disposable personal income growth. The regression includes 6 lags of the independent variables and

6 lags of the tone of media coverage, as well as a 12^{th} seasonal lag to account for the tendency of news coverage to be persistent and for coverage to be seasonal (coverage in, for example, February, of one year is related to that in February of the previous year). Inflation, unemployment and the S&P Index are the only economic indicators available monthly covering the full period of our media data, 1947–2010. After computing changes and accounting for lags, we have 749 observations with which to answer the question: controlling for past coverage—a conservative criterion—which of these economic variables influences the tone of news coverage? Beginning in 1959, real disposable income is also available monthly so we extend our analysis to include 6 lags of this variable as well. Our results suggest that the performance of the stock market is related to the tone of economic news coverage when considering the full period, while changes in unemployment seem related to media tone over the recent 1978-2010 period. Disposable income appears to be unrelated to the tone of news coverage, controlling for the tone of previous coverage.

[**Table 6 Here**]

We also report results from the same analysis over the more recent time period from 1978–2010. We do so for two reasons. First, the media data was trained on human-coded data from 1980 on such that to the extent the nature of news coverage has changed it may be more reliable in the more recent era. Second, our analysis of economic perceptions begins in 1978 when economic perceptions are first available on a monthly basis. In this analysis month to month changes in unemployment (ρ=0.08) are related to the probability an article is positive in tone. Neither inflation nor disposable income add significantly to our ability to predict tone during this period.

5 The Sources of Economic Perceptions: Economic Performance and the Tone of Economic News Coverage

Our next question is: Does news coverage of the economy predict consumer perceptions of the economy controlling for economic conditions themselves? To answer this question we estimate error correction models of the University of Michigan Index of Consumer Sentiment, evaluations of business conditions both one year ahead and five years ahead, and evaluations of business conditions today compared to one year ago.[18] We include our four individual economic indicators—(changes in) unemployment, percent change in the S&P Index, inflation, and (annualized) growth in real disposable personal income—as well as the Conference Board Coincident Indicator Index to measure economic performance. And we include the current period's average probability an article is positive. We lag the economic variables because the *previous* month's value is reported in any given month but we do not lag the tone of media coverage under the assumption that tone over the month is uniform and respondents surveyed about the perceptions are thus exposed to the same tone regardless of when in the month they were surveyed. The alternative is to assume that the previous month's tone influences sentiment, but given what we know about the (short) memory of citizens, it seems to us more likely that recent tone matters in voter evaluations.

[Table 7 Here]

[18]The sentiment measures are at a minimum strongly autoregressive. To ensure stationarity we estimate the model in first differences of our sentiment measures and report the significance of lagged levels of sentiment using the Dickey Fuller critical values. All other independent variables are stationary.

The results of our analysis suggest several things. First, sentiment responds to economic conditions much as we expect. An increase of a tenth of a point in unemployment produced an average decrease in overall consumer sentiment of about a quarter point in the next month. The effect was larger on prospective evaluations, particularly assessments of the national economy one year ahead, where a tenth of a point increase in unemployment led consumers to take a more pessimistic view of the future to the tune of about a three quarters of a point in the next month. The effect was almost as large on retrospective evaluations. Rises and falls in stock market, as measured by the S&P Index growth, were mirrored with changes in consumer sentiment across all four measures of sentiment in the subsequent month. A one standard deviation change of 3.5% in the performance of the S&P has a small but significant effect on sentiment ranging from under a tenth of a point to two tenths. The effects of inflation on consumer sentiment are also largest on one year ahead evaluations of the national economy. Inflation averaged 3.8 points from 1978 to 2010. A one standard deviation (4.1) shift in inflation has an expected effect on these economic perceptions of just under two points (1.76) in the following month. The effect of personal disposable income growth is both substantively and statistically insignificant in these models of economic perceptions. Our final economic indicator, percent change in the coincident economic indicator index, is positively signed and statistically significant. A percentage point increase in the index (the rate of change in the index averages 0.14 with a standard deviation of 0.74 and ranges from -3.3 to 1.89 over this time period) having an expected effect on economic evaluations of from just under a point (0.849 points and 0.898 points) for the overall ICS and five year ahead evaluations of the national economy to just over two points (2.198) for one year head evaluations and 3.455 points for retrospective evaluations, all in the subsequent month.

Once we account for the "objective economy", what independent information does the tone of media coverage of the economy contribute to our knowledge of consumer evaluations? Over this time period our media variable ranges from a 0.28 to 0.61 likelihood of conveying positive tone. The average monthly value is 0.44 with a standard deviation of 0.06. According to our estimates tone contributes no information to *retrospective* evaluations: controlling for economic conditions, the nature of economic news had no statistically significant effect on how people viewed the past. This is perhaps unsurprising, as people have direct experience with the economic past and need not appeal to the media to help them decide what to make of it. Forecasting the future, and this is indeed what is asked of people when assessing the economic future (low stakes though it is), may make people more likely to think about what they've heard or read, to mediate their experience or the explain the meaning of economic statistics. And it is with future-oriented assessments that media coverage exerts the largest effect on economic evaluations. While a standard deviation change in the likelihood a month's stories are positive has an expected effect of just half a point on overall economic evaluations (the ICS), the estimated effect is one point for one year ahead forecasts and just under that for five year ahead forecasts.

6 Conclusion/Future Research

Given the noise of our measure of media-tone, the results here are presented with the caveat that they are quite preliminary. However, our initial attempt at machine-learning based coding of media tone suggests that with a more temporally representative training data set, and perhaps with finer-grained conceptual questions of tone, we should be able to produce accurate measures of media tone for a wide array of media sources. We believe this will allow us to better estimate what aspects of the macro-economy affect media tone, and examine how those effects vary across media sources.

References

Ansolabehere, Stephen, Marc Meredith & Erik Snowberg. 21012. "Mecro-economic Voting: Local Information and Micro-Perceptions of the Macro-Economy." Working Paper.

Arel-Bundock, Vincent. 2014. *countrycode: Convert country names and country codes*. R package version 0.17.
*http://CRAN.R-project.org/package=countrycode

Bartels, Larry M. 2008. *Unequal Democracy: The Political Economy of the New Gilded Age*. Princeton University Press.

Benoit, Kenneth, Drew Conway, Michael Laver & Slava Mikhaylov. 2014. "Crowd-sourced data coding for the social sciences." Unpublished Manuscript.

Blood, Deborah J. & Peter C.B. Phillips. 1995. "Recession Headline News, Consumer Sentiment, the State of the Economy and Presidential Popularity - A Time Series Analysis 1989-1993." *International Journal of Public Opinion Research* 7:1–22.

De Boef, Suzanna & Paul M Kellstedt. 2004. "The political (and economic) origins of consumer confidence." *American Journal of Political Science* 48(4):633–649.

Goidel, Robert K. & Ronald E. Langley. 1995. "Media Coverage of the Economy and Aggregate Economic Evaluations: Uncovering Evidence of Indirect Media Effects." *Political Research Quarterly* 48:313–328.

Hastie, Trevor, Robert Tibshirani, Jerome Friedman, T Hastie, J Friedman & R Tibshirani. 2009. *The elements of statistical learning*. Springer.

Hester, Joe Bob & Rhonda Gibson. 2003. "The Economy and Second-Level Agenda Setting: A Time-Series Analysis of Economic News and Public Opinion About the Economy." *Journalism & Mass Communication Quarterly* 80(1):73–90.

Hetherington, Marc J. 1996. "The Media's Role in Forming Voter's National Economic Evaluations in 1992." *American Journal of Political Science* 40:372–395.

Hibbs, Douglas A. 2012. "Obama's reelection prospects under ?Bread and Peace? voting in the 2012 US presidential election." *PS Political Science and Politics* 45(4):635.

Hopkins, Daniel J & Gary King. 2010. "A method of automated nonparametric content analysis for social science." *American Journal of Political Science* 54(1):229–247.

Key, V.O. 1964. *Politics, Parties, and Pressure Groups.* Crowell.

Key, V.O. 1966. *The Responsible Electorate: Rationality in Presidential Voting 1936-1960.* Cambridge, Mass: Harvard University Press.

Linn, Suzanna & Jonathan Nagler. 2014. "Economic Voting and Economic Inequality." *Midwest Political Science Association, Chicago, Illinois* .

Mutz, Diana C. 1992. "Mass Media and the Depoliticization of Personal Experience." *American Journal of Political Science* 36:438–508.

Mutz, Diance C. 1994. "Contextualizing Personal Experience - The Role of Mass Media." *Journal of Politics* 56:689–914.

Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot & E. Duchesnay. 2011. "Scikit-learn: Machine Learning in Python." *Journal of Machine Learning Research* 12:2825–2830.

Pruitt, Stephen W. & George E. Hoffer. 1989. "Economic News as a Consumer Product: An Analysis of the Effects of Alternative Media Sources on the Formation of Consumer Economic Expectations." *Journal of Consumer Policy* 12(1):59–69.

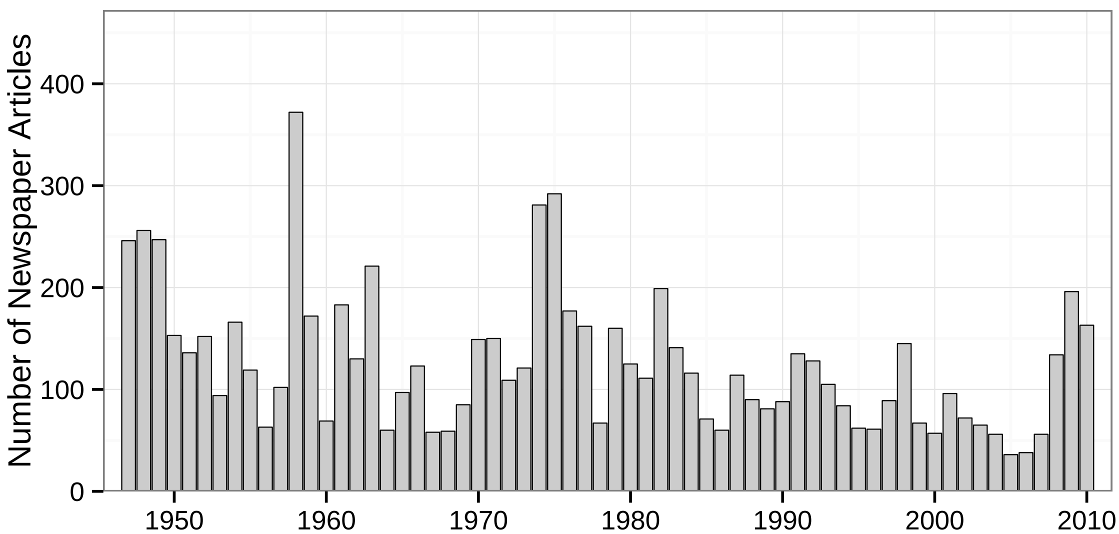
R., Tims Albert, David P. Fan & John R. Freeman. 1989. "The Cultivation of Consumer Confidence: A Longitudinal Analysis of News Media Influence on Consumer Sentiment." *Advances in Consumer Research* 16:758–770.

Sanders, David, David Marsh & Hugh Ward. 1993. "The Electoral Impact of Press Coverage of the British Economy, 1979-1987." *British Journal of Political Science* 23:175–210.

Stevenson, Robert L., William J. Gozenbach & Prabu David. 1994. "Economic Recession and the News." *Mass Communication Review* 21(1-2):4–19.

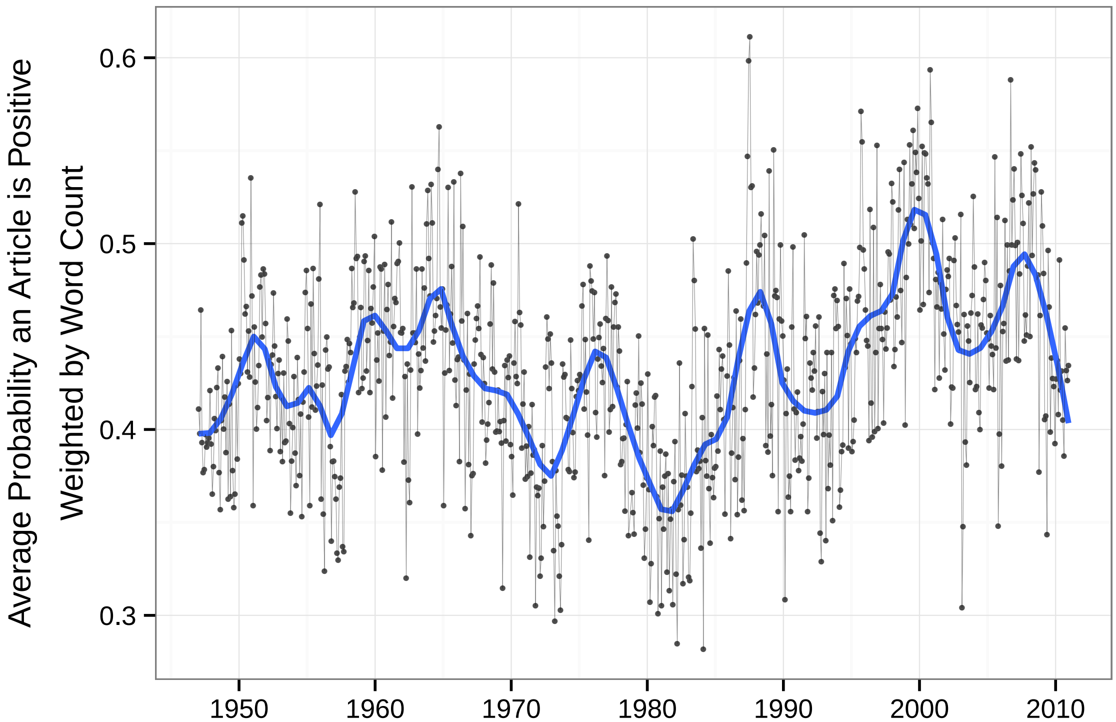
Young, Lori & Stuart Soroka. 2012. "Affective news: The automated coding of sentiment in political texts." *Political Communication* 29(2):205–231.

Figure 1: **Number of Stories per Year**



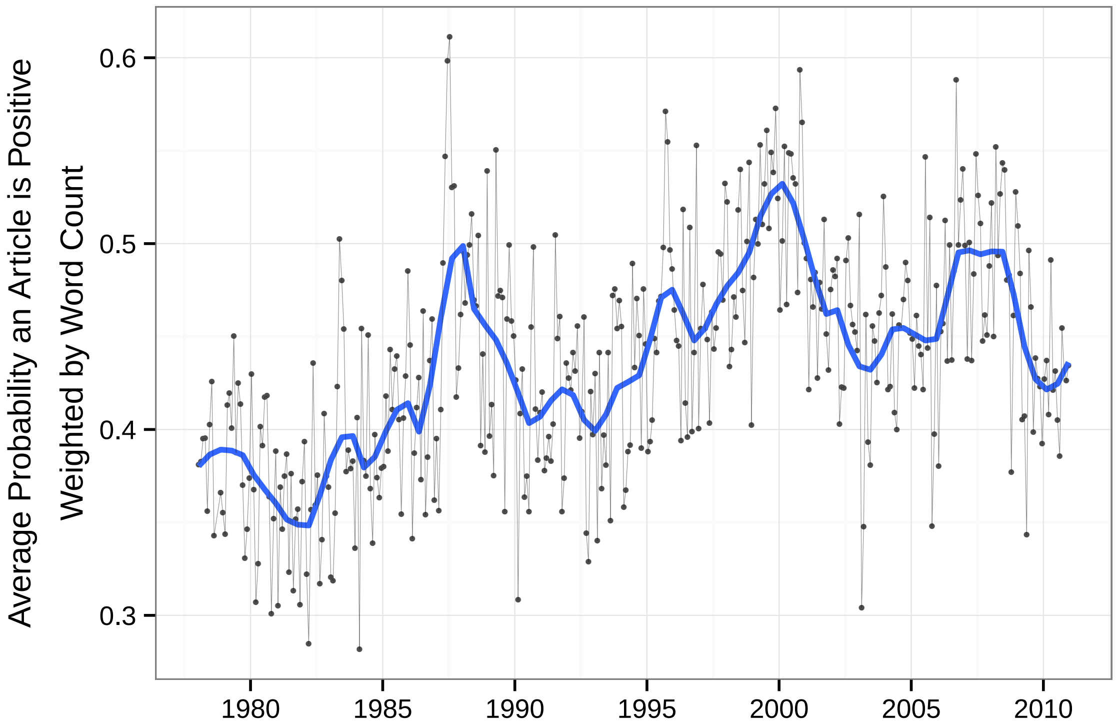
Note: each bar represents the total number of stories included in our dataset of New York Times articles about the economy for each year.

Figure 2: **Sentiment Data, 1947–2010**



Note: each dot represents the average probability that an article published on the New York Times in a given month has a positive tone. The blue line is a loess curve with smoothing parameter $\alpha = 0.10$.

Figure 3: **Sentiment Data, 1978–2010**



Note: each dot represents the average probability that an article published on the New York Times in a given month has a positive tone. The blue line is a loess curve with smoothing parameter $\alpha = 0.10$.

Table 1: Summary of Intercoder Reliability

Relevance

	Undergrads	Crowd (World)	Crowd (US)
Undergrads	**0.82**		
Crowd (World)	0.63	**0.69**	
Crowd (US)	0.34	0.34	**0.68**

Tone

	Undergrads	Crowd (World)	Crowd (US)
Undergrads	**0.82**		
Crowd (World)	0.51	**0.72**	
Crowd (US)	0.52	0.39	**0.86**

Note: values in the diagonals indicate the average pairwise agreement rate within each source. Values off the diagonal indicate percent agreement in modal response to each question across different sources. See Table 2 for more information about the sample size of each dataset.

Table 2: Summary of Classifier Performance

Relevance

Source	N	Acc-uracy	Pre-cision	Bas line
Undergrads	1,410	0.80	0.82	0.65
Crowd (all)	4,000	0.62	0.60	0.54
Crowd (best)	1,999	0.69	0.66	0.61
Crowd (US)	2,000	0.94	0.50	0.93

Tone

Source	N	Acc-uracy	Pre-cision	Bas line
Undergrads	406	0.73	0.55	0.72
Crowd (all)	1,156	0.60	0.55	0.56
Crowd (best)	571	0.59	0.59	0.50
Crowd (US)	95	0.62	0.62	0.57

Note: N indicates sample size; Accuracy indicates 5-fold cross-validated accuracy; Precision indicates 5-fold cross-validated precision; and Baseline indicates baseline (proportion of responses in modal category). The sample size for the tone classifier is smaller because we filter news stories that are classified as not relevant.

Table 3: Confusion Matrix of Machine Learning Classifier

	Training Data	
	Negative	Positive
Pr(positive)<=0.50	405	258
Pr(positive)>0.50	184	809

Table 4: Top Predictive N-grams in Classifier

Negative n-grams
sharp, weak, tuesday, weakness, began, cuts, lost, bad, editor, “growth falls”, falls, stocks, figure, fact, late, worse, business, cut, “rising unemployment”, gas, federal, column, “type letter”, things, unless, disappointing, hit, “labor department said”, unemployed, “jobless benefits”, grim, “rate rose”, release, corporations, recession, reagan, “job losses”, “report showed”, figures, layoffs, abstracts january, slowing, needs, having, forecast, street, flat

Positive n-grams
lowest, strong, businesses, improvement, program, good, stronger, “new jobs”, ended, “growth short”, services, country, gain, steady, finally, lift, gains, proposal, added, credit, “economists expectations”, pressures, “economic stimulus”, buying, continues, benefit, strength, encouraging, “journal abstracts december”, “abstracts december”, expansion, peaked, fell week, lowell, “low inflation”, “million people”, chance, editorial, payroll, high tech, service, greenspan, slightly, “august saturday”, approved, boost, subsidies, mean, better, inflation

Table 5: **Correlations: Media, Real Economy, Economic Perceptions**

Correlations in Levels: 1948 - 2010†, 1959—2010* ($T = 756^{\dagger}$; $T = 623^{*}$)^{a}

	Media Tone	**Article Count**	**Unem ploy ment**	**Infla- tion**	**S&P Index**	**Change in Disp Income**	**Retro- spective Evalu- ations**	**Pro- spective Evalu- ations[b]**	**Lag Index**	**Coin Index**
Article Count†	-0.33									
Unemployment†	-0.19	0.45								
Inflation†	-0.14	0.25	0.04							
Δ S&P Index†	0.09	-0.10	0.10	-0.15						
Δ Disposable Income*	0.01	-0.03	-0.01	-0.18	0.13					
Δ Lagging Index*	0.02	-0.14	-0.33	0.13	-0.23	-0.09				
Δ Coincident Index*	0.02	-0.15	0.16	-0.23	0.43	0.20			-0.24	
ΔLeading Index*	-0.00	-0.14	-0.12	-0.09	0.11	0.32			-0.06	0.63

Correlations in Levels: including Economic Perceptions 1978—2010 ($T = 396$);

	Media Tone	**Article Count**	**Unem ploy ment**	**Infla- tion**	**S&P Index**	**Change in Disp Income**	**Retro- spective Evalu- ations**	**Pro- spective Evalu- ations[b]**	**Lag Index**	**Coin Index**
Article Count	-0.45									
Unemployment	-0.49	0.51								
Inflation	-0.20	0.27	0.00							
Δ S&P Index	0.03	-0.11	0.08	-0.11						
Δ Disposable Income	0.02	-0.01	-0.02	-0.15	0.12					
Bus Retrospections	0.21	-0.23	-0.34	-0.09	0.14	0.10				
Bus Prospections	0.30	-0.29	-0.37	-0.26	0.16	0.14	0.88			
ΔLagging Index	0.06	-0.12	-0.30	0.17	-0.21	-0.08	0.39	0.22	0.13	
ΔCoincident Index	-0.05	-0.07	0.20	-0.11	0.45	0.19	0.41	0.44	-0.25	
ΔLeading Index	-0.03	-0.07	-0.10	-0.01	0.14	0.36	0.56	0.49	-0.01	0.64

[a]Correlations involving the Lagging, Leading, and Coincident Economic Indicator Indices cover the period 1959–2010. Remaining correlations cover the period 1948–2010, inclusive. Indices are measured in growth rate over the previous month.

Economic variables are measured in levels, except for real per capita disposable income, which is measured in changes; and the SP500, which is measured in percentage change from the previous month.

Both retrospective evaluations and prospective evaluations are focused on ‘business conditions’, prospective evaluations are about business conditions 1 year ahead.

Table 6: Models of Average Probability an Article is Positive: Block F-Tests

	(1) 1947–2010 $T = 749$	(2) 1959–2010 $T = 624$	(3) 1959–2010 $T = 617$	(4) 1978–2010 $T = 396$
Inflation	0.17	0.28	0.37	0.15
Changes in Unemployment	0.19	0.24	0.22	0.08
Percent Change in Real S&P	0.02	0.20	0.24	0.66
Real Disposable Income Growth			0.16	0.44

Note: Dependent variable is the average probability an article is positive, measured monthly. Data on the S&P is from Shiller `http://www.econ.yale.edu/~shiller/data.htm` and is measured in percent change from the previous period. The remanning economic data was obtained through FRED. Disposable income growth is measured as (annualized) percent change in billions of 2009 chained dollars. Unemployment is seasonally unadjusted, measured as change from the previous period. Monthly inflation data is (annualized) percent change in the consumer price index for all goods (all urban consumers). Six lags of all economic variables are included in the model. Six lags of the dependent variable and a seasonal lag (12) are also included.

Cell entries are p-values for block F-Tests on the group of lagged measures of each row-variable.

Table 7: Error-Correction Models of Economic Perceptions

	(1) Index of Consumer Sentiment	(2) One Year Prospective Evaluations	(3) Five Year Prospective Evaluations	(4) One Year Retrospective Evaluations
Sentiment_{t-1}	-0.113*	-0.138*	-0.134*	-0.090*
	(0.018)	(0.021)	(0.025)	(0.013)
$\text{Unemployment}\Delta_{t-1}$	-2.552*	-7.880*	-2.720	-6.765*
	(1.252)	(3.378)	(2.385)	(2.925)
$\text{S\&P}\Delta_{t-1}$	0.016*	0.058*	0.025*	0.052*
	(0.005)	(0.014)	(0.010)	(0.012)
Inflation_{t-1}	-0.174*	-0.431*	-0.289*	-0.348*
	(0.049)	(0.132)	(0.097)	(0.111)
$\text{Disposable Income}_{t-1}$	-0.007	-0.028	-0.009	0.025
	(0.021)	(0.057)	(0.041)	(0.048)
Media Tone_{t}	9.028*	16.556+	15.716*	3.243
	(3.380)	(8.897)	(6.650)	(7.456)
Coincident Indicator	0.849*	2.198*	0.898	3.456*
Index Δ_{t-1}	(0.312)	(0.861)	(0.605)	(0.740)
$\text{Sentiment}\Delta_{t-1}$	-0.086+	-0.058	-0.225*	0.032
	(0.051)	(0.051)	(0.050)	(0.049)
$\text{Sentiment}\Delta_{t-2}$	0.118*			
	(0.048)			
Constant	6.386*	8.183*	6.237*	7.631*
	(1.760)	(4.060)	(3.057)	(3.366)
R-squared	0.164	0.161	0.150	0.237
RMSE	3.660	10.039	7.236	8.552
Bruesch Godfrey (12 lags)	0.35	0.35	0.24	0.13

Standard errors in parentheses

+ $p < 0.10$, * $p < 0.05$

Note: Models are of changes in consumer sentiment. The coefficient on lagged sentiment follows a Dickey Fuller distribution. Data on the S&P is from Shiller `http://www.econ.yale.edu/~shiller/data.htm` and is measured in percent change from the previous period. The coincident economic indicator index (CEI) is measured as percent change from the previous period and is from The Conference Board. The remanning economic data was obtained through FRED. Disposable income growth is measured as (annualized) percent change in billions of 2009 chained dollars. Unemployment is seasonally unadjusted, measured as change from the previous period. Monthly inflation data is (annualized) percent change in the consumer price index for all goods (all urban consumers).