The Power of Narrative Sentiment in Economic Forecasts[*]

Steven A. Sharpe
Nitish R. Sinha
Christopher A. Hollrah

First draft: December 3, 2019
Current draft: June 29, 2021

Abstract

The sentiment, or "Tonality", extracted from the narratives that accompany Federal Reserve Board (Greenbook) economic forecasts is strongly correlated with future economic performance, positively with GDP and negatively with unemployment and inflation. More notably, Tonality conveys substantial incremental information, as it predicts errors in both Federal Reserve and private-sector point forecasts of unemployment and GDP growth up to four quarters out, with more favorable sentiment predicting better than expected economic performance. Tonality also has power for positively predicting both monetary policy surprises and stock returns up to four quarters ahead. Quantile regressions indicate that much of Tonality's forecasting power arises from its signal of downside risks to economic performance and stock returns. Moreover, if observed in real time, tonality would have been particularly informative about prospective economic performance and stock returns at times when economic uncertainty was high or when point forecasts called for subpar GDP growth.

JEL codes: E17, E52, G14.

[*] Our views do not necessarily reflect those of the Federal Reserve System nor its Board of Governors. We are very grateful for the research assistance provided by Toby Hollis, Taryn Ohashi, and Stephen Paolillo. Many thanks to Jeremy Rudd for help in developing the wordlists. We are also thankful to our Board colleagues Jack Bao and Neil Ericsson for detailed discussions.

The Power of Narrative Sentiment in Economic Forecasts

Abstract

The sentiment, or "Tonality", extracted from the narratives that accompany Federal Reserve Board (Greenbook) economic forecasts is strongly correlated with future economic performance, positively with GDP and negatively with unemployment and inflation. More notably, Tonality conveys substantial incremental information, as it predicts errors in both Federal Reserve and private-sector point forecasts of unemployment and GDP growth up to four quarters out. Tonality also has power for predicting monetary policy surprises and, even more notably, stock returns up to four quarters ahead. Tonality is most informative about future economic performance and stock returns when economic uncertainty is high and when point forecasts predict subpar growth. Quantile regressions indicate that much of Tonality's forecasting power arises from its signal of downside risks to economic performance and stock returns.

JEL codes: E17, E52, G14.

I. Introduction

Over the years, even as many researchers and market participants have questioned the value of macroeconomic forecasts, substantial resources continue to be devoted to their production and dissemination. For instance, the Blue Chip Survey of Economic Indicators collects monthly forecasts of the U.S. economy from over 50 "top analysts," most of whom are associated with private-sector profit-driven firms. The Blue Chip Financial Forecasts survey polls a similar set of analysts on their interest rate and currency value forecasts, despite probably even less compelling evidence of success in predicting financial prices. Similarly, eight times a year, prior to each meeting of the FOMC committee, the staff at the Federal Reserve Board provide a detailed forecast of the U.S. economy (staff forecast). Our study provides a new perspective on the information embedded in macroeconomic forecasts and their potential value to policymakers and financial market participants.

In the academic literature, macroeconomic forecasts have been evaluated for their predictive content, for evidence of bias, as well as for their comparative merit.[1] Such studies focus almost exclusively on the track record of quantitative point forecasts, usually of inflation and/or GDP growth. Consequently, they largely ignore the narratives that accompany the quantitative forecasts, even though narratives are often a substantial part of the forecasters' product. Such narratives tend to give a flavor of the range of plausible outcomes or characterize the direction of likely risks to forecasts. While difficult to verify, it seems quite plausible that policymakers and investors who pay for these forecasts place significant value on "consuming" the narratives that accompany the quantitative point forecasts.

This study breaks new ground by applying tools from the emerging literature on textual analysis to gauge the incremental forecasting signal of the sentiment extracted from the narratives that accompany forecasts. To do so, we focus on Federal Reserve Board forecasts, which are described in the Greenbook and are perhaps the longest available time series of macroeconomic forecasts for the U.S. economy. We begin by quantifying the degree of

[1] For example, Romer and Romer (2000) show the Federal Reserve Greenbook forecasts are superior to private sector forecasts. D'Agostino and Whelan (2008), Gamber and Smith (2009), Sinclair, Joutz and Stekler (2010) note that the superiority of Fed's forecast has faded recently.

optimism versus pessimism embedded in each forecast narrative, which we call the "Tonality" of the forecast narrative, based upon counts of words classified as positive or negative. The starting point for that classification is the Harvard Psycho-social dictionary, which is then fine-tuned by excluding words that have special meaning in an economic forecasting as well as Federal Reserve context, such as "fed" and "interest."[2] The measure of forecast narrative sentiment that we extract is quite strongly correlated with the strength of the accompanying point forecasts for key economic variables, usually with the intuitive sign. In particular, Tonality is positively correlated with forecasts of GDP growth and negatively correlated with both the unemployment and inflation forecasts.

The central question we consider is to what extent, and why, such a measure of text sentiment might have value as a signal of future economic performance. We examine whether Tonality has incremental power, over and above the point forecasts, for predicting key macroeconomic quantities—namely unemployment, GDP growth, and inflation. We pursue the hypothesis that positive (negative) sentiment predicts more (less) favorable economic outcomes, such as higher (lower) GDP growth, even conditional on the point forecast. For instance, we estimate both ordinary least square (OLS) and quantile regressions that project Greenbook forecast errors on Tonality.

In OLS regressions, we find that Tonality has significant predictive power for both GDP growth and the change in the unemployment rate, a result that holds for forecast horizons from one to four quarters ahead. More positive sentiment in the forecast narrative text predicts higher-than-forecast GDP growth and a lower-than-forecast unemployment rate. One implication of these results is that Greenbook point forecasts are not "efficient", that is, mean squared forecast errors could have been smaller if the point forecasts had incorporated all the information embedded in the forecast narrative.

What is more, the quantile regressions reveal strong asymmetry in Tonality's predictive content. For both GDP growth and the unemployment rate, and at all horizons, the results indicate that Tonality is most informative about the likelihood of bad economic news; that is, it provides a particularly strong signal of lower tail risks for GDP growth, relative to forecast, and

[2] For instance,"fed" is a negative word in the Harvard dictionary as it is often used commonly as "fed up."

signals the prominence of upper tail risks for unemployment relative to forecast. While Tonality does not exhibit a clear directional signal for inflation, lower Tonality does appear to signal larger tail risks to inflation forecasts in either direction.

The asymmetry of Tonality's predictive content for GDP growth has notable parallels to some previous findings where quantile regressions are used to predict macroeconomic risks. In particular, both Hengge (2019) and Rogers and Xu (2019) find that high economic uncertainty predicts larger downside risks to future GDP growth, while conveying little information about mean outcomes. Similarly, Adrian, et al. (2019) find that a financial conditions index has substantial predictive power for the extent of negative tail risk to GDP growth but relatively little predictive power for mean or median GDP growth. While we show that the signal from Tonality has some commonality with that conveyed by both uncertainty and financial conditions measures, we find that Tonality has marginal predictive power even after controlling for those factors.

One possible explanation for Tonality's predictive power could be stickiness in the Greenbook point forecasts, that is, forecast revisions that are more sluggish than would be optimal for minimizing mean square error. Nordhaus (1987) first described this as "Inefficient forecasts … let the news seep in slowly" and argues that the resultant forecast errors would be predictable, in part, using recent forecast revisions.[3] Such an inefficiency in Greenbook point forecasts could account for the predictive power of the narrative if the sentiment in the narrative is simply more "nimble" to incorporate new information. We find little evidence that sticky forecasts can account for much of Tonality's predictive power.

To get a better sense of the nature of the information conveyed by Tonality, and when it might be most useful, we examine whether its predictive power is stronger when macroeconomic uncertainty is high or when the GDP forecast calls for below-trend growth. One long-perceived weakness of economic forecasts, documented early on by Zarnowitz and Braun (1993) and revisited recently by Smirnov and Avdeeva (2016), is that point forecasts rarely call for an

[3] More recently, in an analysis of consensus forecasts from the Survey of Profession Forecasters, Coibion, and Gorodnichenko (2015) find evidence of "information rigidity," in that forecast revisions for inflation tend to predict future forecast errors in the same direction. Andrade and Le Bihan (2013) as well as Dovern, et al. (2015) go even further, showing that individual forecast revisions also tend to predict an individual forecasters errors in the same direction, though the magnitude of rigidity is smaller than in consensus forecasts.

outright decline in GDP before a recession has actually begun. Thus, a plausible hypothesis is that the information value of Tonality is higher when the four-quarter outlook calls for subpar growth; at such times, the forecast narrative might be particularly informative about the balance of risks surrounding the outlook. Indeed, we find that, either when uncertainty is high or when the four-quarter GDP forecast calls for sub-par growth, the predictive power of point forecasts is especially poor; at the same time, Tonality conveys a sizable amount of information about the likely direction of the forecast error.

To examine the informational value of Tonality for economic agents outside the Fed, had they observed it in real time, we merge our data on Greenbook Tonality together with roughly contemporaneous consensus economic forecasts compiled in Blue Chip Financial Forecasts and find that Tonality has very similar power to predict errors in the Blue chip forecasts. Here again, the predictive power of Tonality for economic activity appears to be strongest when the consensus forecast calls for below-trend GDP growth. The similar complementarity of Tonality with private sector forecasts indicates that the information content of Tonality is not simply the consequence of some internal Fed forecasting dynamic; rather, the sentiment reflected in the Greenbook narrative would appear to have similar incremental value for consumers of private sector forecasts.

In light of the predictive power of Tonality for economic activity (GDP and the unemployment rate) relative to private-sector forecasts, we consider a logical corollary: does Tonality of the text help to predict monetary policy surprises? If forecasters produce Fed Funds forecasts that are consistent with their point forecasts for the unemployment rate, as dictated by some Taylor-like rule, then, all else the same, one might expect upside surprises to the unemployment rate predicted by Tonality to be accompanied by upside surprises to the fed funds forecast. Indeed, we find that Tonality does have significant predictive power for policy rates relative to Blue Chip forecasts; in particular, a more optimistic tone in the Greenbook text presages a higher than anticipated Fed funds rate up to four quarters ahead.

Finally, we examine whether Tonality predicts stock market returns over horizons similar to those in its predictions of economic forecast errors. This amounts to a test of whether the information embedded in Tonality, had it been publicly available, could have conveyed valuable information for investors. We find that Tonality has substantial power for predicting excess returns on stocks over holding periods that range from the 3 to 12 months subsequent to the completion of the Greenbook forecast. The positive coefficient on Tonality is consistent with the view that, on average, the beneficial cash flow and investor risk premium effects far outweigh any negative implications from tighter monetary policy.

Of course, unlike the standard conditioning variables used in the return predictability literature, Tonality is not directly observable to investors; but this begs the question of whether policymakers indirectly convey the information in Greenbook Tonality to the public. This question is closely related to the burgeoning literature on information signaled by the words in FOMC communications, which we consider to be largely beyond the scope of this paper. Nonetheless, before concluding, we briefly examine whether the sentiment gauged by Greenbook Tonality is reflected in the two subsequent formal FOMC communications that follow, the FOMC statement released at the end of the meeting and the FOMC meeting minutes released several weeks hence. We find that the Tonality of the relatively terse FOMC statements has low correlation; in contrast, Tonality measured from the FOMC minutes correlates noticeably with the recent Greenbook's Tonality. This suggests a promising direction for future research on the information conveyed to the market in the FOMC minutes.

Section II provides an overview of how the paper relates to some existing lines of research. Section III describes how we measure Tonality and explores how it co-varies with the point forecasts of key macroeconomic variables in the Greenbook. In section IV, we examine the extent to which Tonality conveys information about future macroeconomic conditions not already reflected in point forecasts. Section V examines the relevance of the information in Tonality for market participants, specifically, its ability to predict errors in the Blue Chip consensus forecasts and its ability to predict future monetary policy surprises and stock returns. Finally, it briefly examines whether Greenbook Tonality is transmitted to the public in either the post-meeting FOMC statements or the FOMC meeting minutes. Section VI concludes.

II. Related Literature

This study contributes to the literature on the efficacy of economic forecasts by providing a broader perspective on the nature of economic forecasts beyond what can be learned by studying the properties of forecast errors in isolation. A singular focus on numerical point forecasts ignores other quite pertinent information that forecasters bring to the table but are not reflected in those forecasts. In addition, our study also contributes to the relatively new and burgeoning line of research in economics that draws insights from treating text as a new source of data. Thematically, our work is related to the nascent research in economics and finance that attempts to quantify narratives, an agenda recently nudged into the mainstream, in part, by Shiller's (2017) presidential address to the American Economic Association.

Our study also echoes elements of those that examine whether the tone of newspaper articles helps explain or predict stock market returns, beginning with Tetlock (2007), using techniques elaborated upon more recently, for instance, by Garcia (2013), Heston and Sinha (2017) , Calomiris and Mamaysky (2019) and Ke, Kelly and Xiu (2019). It also is related to Asquith, Mikhail and Au (2005), which examines whether the sentiment of the text in Wall Street analyst reports explains firms' stock price responses to earnings forecast revisions. Perhaps the study closest in purpose to ours is that by Clements and Reade (2020), which measures the sentiment of the narratives in the Bank of England Quarterly Inflation Reports and finds that sentiment can help predict errors in point forecasts for output growth one to two quarters ahead.[4]

Also related are recent studies that quantify information conveyed in monetary policy communications and by gauging its impact on market prices. Stekler and Symington (2016) and Ericsson (2016) manually score the sentiment reflected in the FOMC minutes and find that it maps quite closely to the current-quarter and quarter-ahead GDP forecasts in the correspond Greenbook. Hansen and McMahon (2016) parse FOMC statements into the information conveyed about either forward guidance or economic conditions and find that the forward guidance has more noticeable market impact. Hansen and McMahon (2017) use text analysis to infer change in the nature of FOMC deliberation following increased transparency. Schmeling and Wagner (2017) gauge the tone of European Central Bank press conferences and find that a more positive tone induces higher interest rates and lower credit spreads and equity volatility. Carvalho, Hsu and Nechio (2016) use sentiment quantified from FOMC communications to examine interest rate reactions to FOMC communication during the zero lower bound period.

---

[4] Similarly, but using a manual scoring approach, Jones, Sinclair and Stekler (2019) quantify in those Bank of England reports the narratives related to inflation and find their metric helps predict quarter-ahead inflation.

Our study differs from these in that we focus on sentiment embedded in the communications between Fed staff and the FOMC committee, information that is only available to the public years later.

The paper also speaks to the measurement of time-varying macroeconomic uncertainty. For instance, the approach popularized by Baker, Bloom and Davis (2016), counts uncertainty related words in newspapers. Jurado, Ludvigson and Ng (2015) propose to measure more directly macroeconomic uncertainty as the conditional volatility of the unforecastable component of macroeconomic variables based on all available data, including financial variables. Relatedly, Clark, McCracken and Mertens (2020) construct an uncertainty measure using the forecast errors of macroeconomic forecasters. We find that the tone of forecast narratives conveys substantial information over and above popular uncertainty measures. At the same time, we find that the forecast narrative is most informative when macroeconomic uncertainty is high.

III. Measurement of Tonality in Greenbook Text

A. Measuring Tonality

Prior to every scheduled FOMC meeting, Federal Reserve Board staff puts together its forecast for the U.S. economy in an internal Fed document called the *Greenbook* (now the *Tealbook*), which is only made public after a 5-year lag. The Greenbook (and subsequently Tealbook) contains both a point forecast of macroeconomic variables accompanied by a narrative describing and explaining the point forecast, as well as a characterization of recent economic and financial developments. The forecast and accompanying narrative are hammered out in tandem over roughly a two-week week period, and the narrative is usually finalized with a day or two after the point forecast is "closed".

Greenbook forecasts were produced monthly until 1981; thereafter, their frequency dropped to eight per year. Our sample begins January 1970, shortly after the staff's quantitative quarterly forecast began to look forward more than two quarters. Beginning in August 1974, text analysis is conducted on the text of Greenbook Part 1, the Summary and Outlook, the section of the document that focuses on the forecast. Prior to the division of Greenbook into Parts 1 and 2, our analysis uses the text from the section titled Recent Developments and Outlook for Domestic Economic Activity. Our sample ends in December 2009, shortly before Greenbook was replaced by Tealbook A, when Greenbook was consolidated with closely related material from the (also retired) Bluebook.

We construct an index that quantifies the optimism and pessimism of the Greenbook text, which we call "Tonality." Tonality is constructed as a weighted sum of positive words minus a weighted sum of negative words. To classify words as "positive" or "negative," we use for our starting point the Harvard IV Psychosocial Dictionary, a general dictionary of written English that identifies words having either positive or negative connotations. We construct the intersection of those word lists with the words that appear at least once in our full set of documents. Positive words in our dictionary include terms like "enthusiasm," "abundant," "enhance," and "successful," whereas examples of negative words include "unrest," "fragile," "trouble," and "gloomy." Because some words that have a negative (or positive) connotation in conversational English do not normally have a negative (positive) connotation in an economic forecasting, we cull these lists to create a custom dictionary of 231 positive words and 102 negative words that appear in our documents.[5] For example, in contrast to the psycho-social dictionary, the words "demean" or "hedge" would not be considered as negative in an economic outlook context.

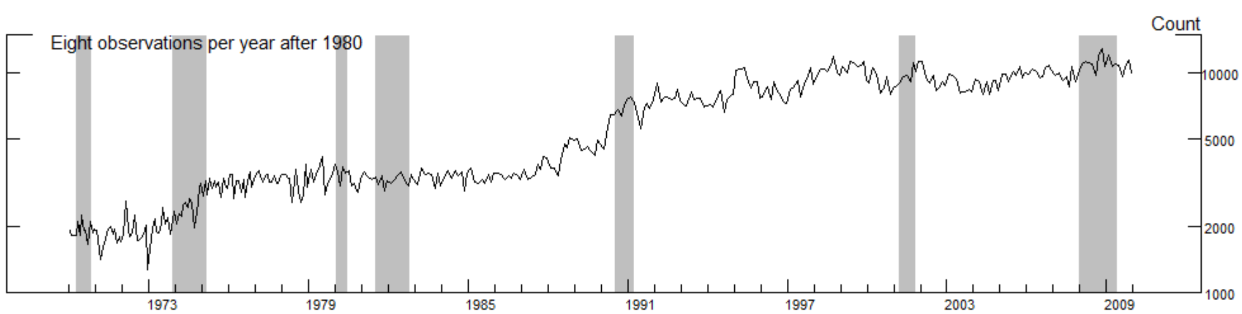
An alternative option might have been to use the Loughran and McDonald (LM) dictionary, which was also constructed starting with the Harvard dictionary, but they cull out words that have a different connotation in accounting documents (10Ks) issued annually by public firms. For instance, the LM dictionary takes out words such as "liability," a common balance sheet term, and "oversight" which in the 10-K often refers to as "someone with an oversight of a business area" rather than a "mistake in analysis." Seeing as the LM dictionary is tuned to accounting and individual firm level discussion, we are essentially following its authors general approach of using domain-specific knowledge to customize, rather than simply adopting their dictionary, which was customized for their domain.

[5] For the list of positive and negative words, see the online publication appendix A.

Our approach to word counting is similar to Loughran and McDonald (2011) and other older studies such as Tetlock (2007), in that they also examine word frequency without trying to gauge the context in which words are used. By using the whole document to quantify the overall degree of optimism, irrespective of how words are grouped, we have chosen not to use more elaborate methods of text analysis that would, for instance, attempt to connect the words that convey sentiment with their antecedents, such as particular topics or economic indicators.[6]

While these more recent methods have the potential benefit of providing more context for the role of sentiment, they require a good deal of pre-processing and additional judgment, for instance, on what subjects to focus on or how to group them, or how to classify "nearby" words in text space. It would presumably also exclude a lot of information such as the descriptors of the many other economic variables that are related to the specific indicators on which we focus. In some cases, Greenbook narrative speaks to inflation and unemployment in the same sentence, but how to apportion the narrative sentiment in that sentence might be easy for a human reader to grasp but hard for a program to decipher. As the first study to analyze the sentiment that accompanies the Fed staff forecast, we chose to begin by studying the information content of overall sentiment.

Figure I: Total words in the Greenbook

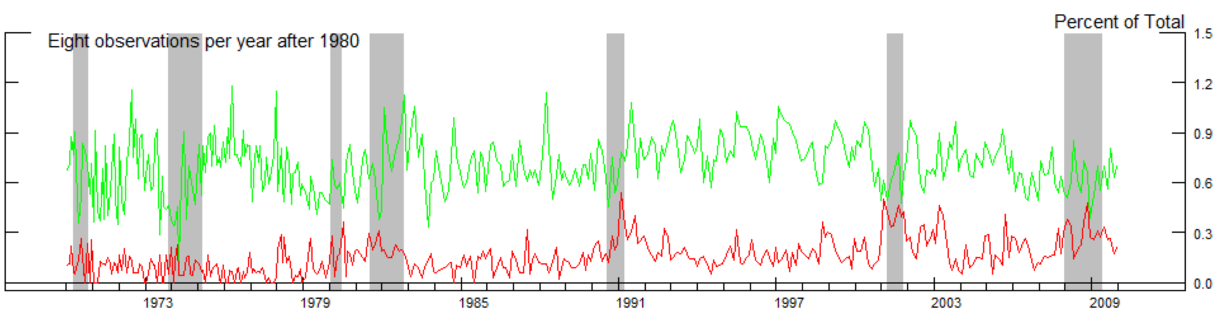


Note: Shaded regions represent NBER-dated recessions. Prior to 1981, Greenbooks were produced nearly every month, thereafter the frequency was reduced to eight times a year.

[6] As one robustness check, we examined sensitivity of our scores to presence of signed words that follow negations. For example, in the clause "GNP is likely to show no further rise", "rise" follows "no" and should not be counted as a positive word. To examine this, we mute all words in a clause that follow words indicating negation using negation word list (no, never, not, nowhere, none) of Das and Chen (2007). The resulting negation-adjusted Tonality measure has a 98 percent correlation with our Tonality measure.

Figure I shows the time series of the total word counts from Greenbook Part I (or its pre-August 1974 equivalent) for our entire sample period. As shown, in the earlier forecast documents, the word count from the outlook section ran at only about 2000 words. After the restructuring in August 1974, the count quickly moved up to about 3000 words, where it hovered until 1990, after which the document gradually ramped up to about 9000 words. **Figure II** shows the number of positive and negative words as a percent of the total word count in each Greenbook. In most documents, the frequency of positive words is far above that for negative words. Also apparent from this picture, prior to the August 1974 restructuring, the percentage of positive words per document appears to have been considerably more variable from one document to the next.

Figure II: Proportion of Positive and Negative Words in the Greenbook



Note: Shaded regions represent NBER-dated recessions. Prior to 1981, Greenbooks were produced nearly every month, thereafter the frequency was reduced to eight times a year. The green line shows the positive words as a proportion of total number of words in that Greenbook. The red line shows negative words as a proportion of total words. Proportions are expressed as percentages.

The Tonality index value of a document compares the number of positive and negative words in its text, using a weighting scheme in which a word's frequency of appearance in any given Greenbook is normalized by its average frequency in other Greenbooks, a weighting scheme commonly known as term frequency-inverse document frequency (tf-idf).[7] Specifically, the weight for each word is equal to its current-document frequency (tf) multiplied by the inverse of its other-documents frequency (idf). For all but the first 40 documents in our sample, we use the *previous* 40 Greenbooks as the corpus for obtaining the idf values for a given Greenbook.

[7] In the information retrieval and text analysis literature the tf-idf weighing scheme is a commonly used metric to gauge the importance of a word in a collection of documents (or a corpus). Loughran and McDonald (2011) first used tf-idf weight in the finance literature to quantify SEC filings by U.S. firms.

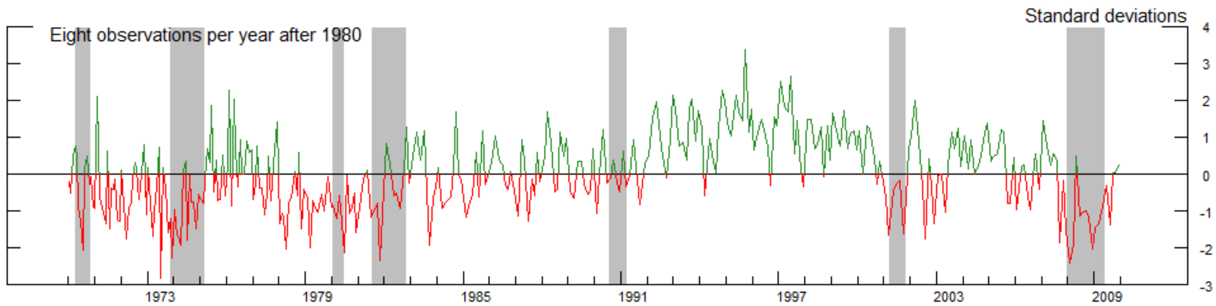
For each of the first 40 documents in our sample, the corpus used for normalization is simply the first 40 documents.[8]

The tf-idf weighing scheme is based on the intuition that infrequently used words are especially informative and so receive relatively high weight in the index, whereas very frequently used words are discounted. Common application of tf-idf scheme would have used the inverse document frequency over *all* the Greenbooks. We chose a moving window of roughly five years to account for changes over time in Greenbook writing style. Nevertheless, the correlation between 40-greenbook rolling window tf-idf scores and a simple tf-idf scheme that "sees" all greenbooks is over 95 percent, suggesting the choice of window does not have a substantial effect on our measure of Tonality. Finally, the Tonality index is standardized to have zero mean and standard deviation equal to one. We adapt the Python machine learning library Scikit-learn Pedregosa, et al. (2011) for tf-idf scoring of Greenbooks.

Figure III shows the Tonality index plotted over the full sample period, with positive (above average) sentiment levels indicated in green and negative levels indicated in red. As one might expect, Tonality appears to be procyclical, with the large majority of observations during recessions being in negative (below average) territory, and a mixture of positive and negative observations during expansionary periods. Among the most deeply negative readings of Tonality are observations in the year leading up to and during the Great recession as well as the 1974-75 recession. The most noticeable run of highly positive readings was during the mid-1990s. Despite these cyclical tendencies, Tonality also appears to be quite volatile, exhibiting spikes that are often quickly reversed. An important consideration is whether these high frequency movements reflect noise in our measure of sentiment or informative innovations.

Figure III: Greenbook Tonality plotted over time

[8] In addition, we treat the set of documents prior to August 1974 as a separate corpus, not necessarily comparable to the later documents; thus, we use solely pre-August 1974 set of documents for measuring the inverse document frequency for these early documents, and similarly for the post-August 1974 set of documents.



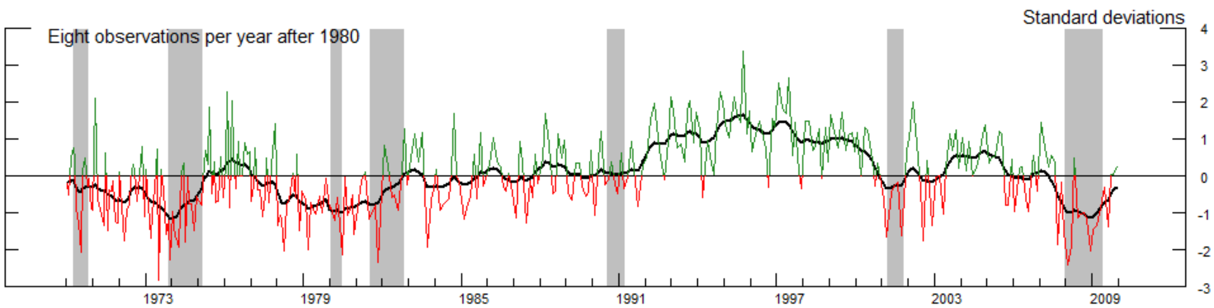
Note: Shaded regions represent NBER-dated recessions. Tonality is standardized to have a zero mean and a standard deviation equal to one. Tonality is shown in green when it is positive and in green when negative. Prior to 1981, Greenbooks were produced nearly every month, thereafter the frequency was reduced to eight times a year.

If the sentiment reflected in Tonality contains information about the perceived state of the economy going forward, then it would seem plausible that the underlying lower-frequency trend in Tonality might contain the lion's share of information of that information. We can think of two reasons that the higher-frequency movements in Tonality may convey less information about underlying sentiment. First, at the individual document level, may be noise created by the fact that the authors contributing to Greenbook change from one forecast round to the next, and authors might contribute idiosyncratic elements to the language used in the text. Perhaps even more important, while an important element of forecast characterization focuses on the underlying trend, a nontrivial portion of the words may be more about characterizing *revisions* to the forecast from the previous Greenbook. Such revisions might be negative even when the underlying trend is positive, and vice versa.

In order to explore the information value of the high- and low-frequency components of Tonality separately, we consider an unobserved components model of the level of Tonality. Moreover, since the macroeconomic literature is littered with evidence that volatility of economic variables changes over time, we use an unobserved components model with stochastic volatility (UCSV). To the extent that there is variation over time in the volatility of inflation (as in Stock and Watson ((2007)) or in economic growth (giving rise to the "great moderation" view), one might naturally expect the same dynamics in the forecast narrative sentiment. From this model, we extract a smoothed measure of Tonality that reflects low frequency movements plus a residual. We call the smoothed component of Tonality, Trend Tonality, while we call the residuals, Tonality shocks.[9]

Figure IV shows the resulting times series plot for Trend Tonality, along with (raw) Tonality. The cyclical pattern in this smoothed measure of sentiment stands out a bit more clearly. Consistent with this interpretation of the residual (Tonality minus trend-Tonality) as Tonality shocks, the autocorrelation coefficient for the residual is close to zero (0.04).

Figure IV: Greenbook Tonality and Trend Tonality plotted over time



Note: The black line shows the time series for Trend Tonality. Shaded regions represent NBER-dated recessions. Tonality is standardized to have a zero mean and a standard deviation equal to one. Prior to 1981, Greenbooks were produced nearly every month, thereafter the frequency was reduced to eight times a year. Tonality is shown in green when positive and in red when negative.

Word clouds depicting the 50 most prominently used positive and negative words in Greenbook during two different time spans (1994-1999, 2005-2009) are shown in **Figure V**. The upper panel shows two side-by-side word clouds for the 50 most prominent positive words in Greenbooks during the two periods. Word size is proportional to its contribution to Tonality, that is, its contribution to the sum of tf-idf weights during the five-year window. Overall, the positive word cloud is a bit bigger during the later period. The substantial overlap in influential words during these two periods suggests little drift over time in the sentiment-related language used. The most important positive word in both periods is "upward", followed closely by

[9] Because the pre-1980 sample used Greenbook forecasts that were produced monthly, rather than eight times per year, we estimate the model parameters for the pre-1981 subsample separately. In previous versions of the paper, trend Tonality was measured as an exponentially weighted moving average (EWMA) of current and past Tonality, which is consistent with a trend produced by an unobserved components model with constant volatility. Following the suggestions of an anonymous referee, we estimated the stochastic volatility (UCSV) model and reject the hypothesis of constant volatility. Even so, the resulting measure of trend Tonality is highly correlated with the previous version based on theEWMA.

"positive." On the other hand, the words "favorable" and "moderation" are more prominent during 1994-1998.

Figure V: Word cloud for fifty most positive and negative words in the Greenbook.

Note: The word cloud on the plot on left side shows fifty positive words frequently used in the Greenbook during the period Jan 1994 through Dec 1998. The word cloud on the right side shows the same for the period Jan 2005 through Dec 2009. The size of individual word in a word cloud is proportional to its contribution in the calculation of Tonality during the plotted time-window.

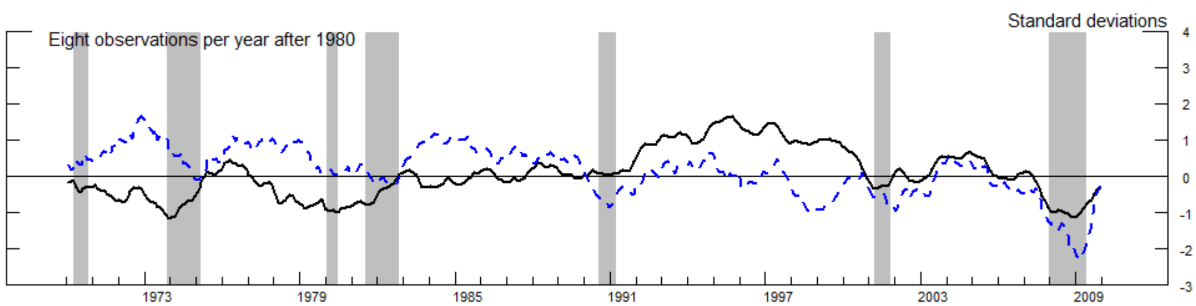
Note: The word cloud on the plot on left side shows fifty most frequently used negative words in the Greenbook during the period Jan 1994 through Dec 1998. The word cloud on the right side shows fifty most negative words during the period Jan 2005 through Dec 2009. The size of a word is proportional to its contribution in the calculation of Tonality during the plotted time-window.

The lower panel shows two side-by-side word clouds for the 50 most prominent negative words in Greenbooks during the same two periods. The most prominent negative word in both samples is, indeed, "negative", followed by "sluggish." Overall, negative words are more prominent in the later period as indicated by the larger word sizes in that cloud. For example, the words "adverse" and "sluggish" are more prominent in 2005-2009 period.

B. Comparing Greenbook Tonality with Alternative Sentiment Measures

While there has been a proliferation in choices for dictionaries used to gauge sentiment in economic and financial research one of the early approaches that remains fairly popular is the classification introduced by Loughran and McDonald (2011). Thus, readers might thus be interested in how Tonality compares with a measure based of Greenbook sentiment built using their dictionary. The most commonly used sentiment measure from Loughran and McDonald's work gauges the prevalence of negative sentiment words only. For purposes of comparison, however, we use the LM word classification to gauge net sentiment (LM-Net), that is, the prevalence of positive sentiment minus the prevalence of negative sentiment, analogous to our construction of Tonality.

Figure VI: Trend Tonality and LM-Net plotted over time

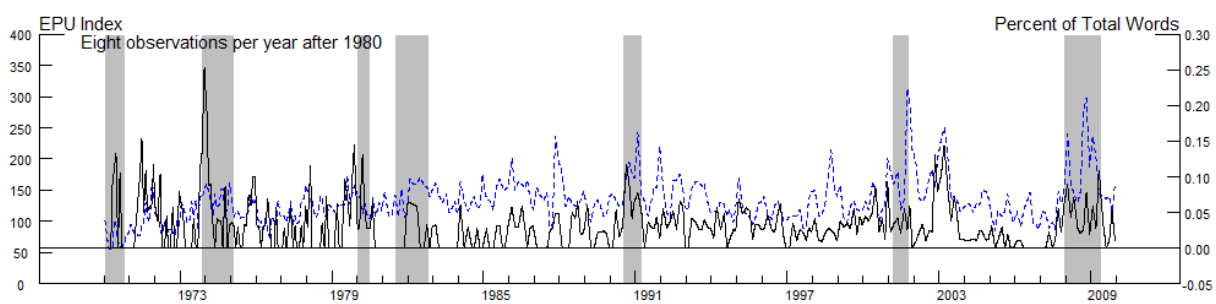


Note: Shaded regions represent NBER-dated recessions. Prior to 1981, Greenbooks were produced nearly every month, thereafter the frequency was reduced to eight times a year. The black line shows Trend Tonality and the dashed blue line is the analogous smoothed measure built from the Loughran and McDonald (2011) dictionary (LM-Net).

In brief, we find that, over the full sample, the correlation of Tonality with LM-Net is quite modest, at 16 percent; not surprisingly, Tonality's correlation with the standard LM measure that uses only negative word counts- is even lower, at 6 percent. A visual sense of how the low-frequency movements in LM-Net compare to those of Tonality is shown in **Figure VI**, which plots Trend Tonality along with a similarly smoothed measure of LM-Net. Broadly speaking, the two series do not appear to have any consistent relationship. In particular, over sizable stretches of the sample, such as the early 1970s and over the 1990s, the two series exhibit a strong negative correlation, followed by long stretches of fairly high positive correlation.

Another fairly widely used metric of sentiment extracted from narrative text in the literature, also relevant to our context, is measures economic uncertainty. Among the most prominent of these comes from the widely cited study by Baker, Bloom and Davis (2016). They construct a measure of uncertainty expressed in news articles, named the Economic Policy Uncertainty index, or EPU. The EPU gauges the relative frequency of "uncertainty" mentions alongside key words that invoke economic, monetary policy, or government policy uncertainty. Because the Greenbook, particularly the section we analyze, consists entirely of economic commentary, our adaptation simply involves counting mentions of "uncertainty" and "uncertain" as a fraction of total word count in Part I of each Greenbook. The resulting measure is plotted in **Figure VII**, alongside the time series for EPU. Notably, in a substantial fraction of the documents, there are no mentions of uncertainty; and, in particular, there are very few mentions of uncertainty even in the run-up to the 2008-2009 financial crisis.[10]

Figure VII: Greenbook Uncertainty plotted over time



Note: Shaded regions represent NBER-dated recessions. Prior to 1981, Greenbooks were produced nearly every month, thereafter the frequency was reduced to eight times a year. Instances of 'Uncertain' and 'Uncertainty' are used to create count of uncertain words, shown as percent of total words (black line), the blue line shows the Baker-Bloom-Davis Economic Policy Uncertainty (EPU) index.

C. Correlation of Tonality with Greenbook Point Forecasts and other Factors

[10] Towards the end of our sample, Federal Reserve staff added a separate "Risk and Uncertainty" section to the Greenbook.

To begin examining the informational content of our text sentiment measures, we calculate their correlations with numerical point forecasts from the associated Greenbook. In particular, we examine simple correlations between Tonality and point forecasts for three key economic performance variables: inflation, the unemployment rate, and GDP growth. The first two are the components of the Fed's "dual mandate" and so of obvious importance. The third, GDP growth, is perhaps the most frequently cited summary statistic of economic performance, and properties of GDP forecasts have received ample attention in the literature. We gauge the forecast of each economic variable for the quarter ahead and over the four-quarter horizon: for the latter, we measure the forecast of cumulative inflation, cumulative GDP growth, and the change in the unemployment rate, each over the subsequent four quarters out (with the current-quarter forecast as the base). We also correlate Tonality with the revisions to the four-quarter forecasts relative to the previous Greenbook. Finally, to gauge the perceived state, or level, of economy activity at the time of the forecast, we use the current-quarter forecast of the unemployment rate.[11]

The correlations of Tonality and its two subcomponents with four-quarter forecasts, and with revisions to those forecasts are shown in the top section of Table 1. The first clear takeaway is that many of the correlations between the forecast measures with Tonality and Trend Tonality are fairly strong, and their signs accord with intuition. Forecast narrative sentiment as measured by Tonality is positively related to expected GDP growth and negatively related to both expected changes in unemployment and expected inflation. Interestingly, all three 4-quarter economic forecasts, as well as current-quarter unemployment, are at least qualitatively more strongly correlated with Trend Tonality than they are with raw Tonality. In contrast, correlations of with *revisions* to the GDP and unemployment forecasts are of somewhat smaller magnitude for Trend Tonality; and only forecast revisions (for unemployment and GDP) are correlated with Tonality shocks, again with the intuitive sign. This suggests that some of the variation in

[11] 4-quarter revisions are measured as changes to the outlook only 3 quarters out. For most observations, constructing revisions to the 4-quarter outlook would require having the lagged value of the 5-quarter outlook, which is frequently unavailable.

Tonality is driven by the direction of revisions from the previous forecast, and that "influence" appears to be reflected in the residual, or shock, component of Tonality.[12]

The lower section of Table 1 shows correlations of Tonality with variables found in other recent studies to help predict economic performance or forecast errors. As shown, Tonality displays only a small negative correlation with EPU-BBD, the Baker-Bloom-Davis measure of uncertainty expressed in newspaper text over the inter-Greenbook period, and similarly with EPU-Gbk, the Greenbook uncertainty as gauged by applying the analogous methodology to the Greenbook text. In contrast, Tonality and Trend Tonality even more so, display a strong negative correlation with both MacroUnc, the economic uncertainty measure devised by Jurado, et al. (2015), and NFCI, the financial conditions index published by the Federal Reserve Bank of Chicago. Lastly, we find that S&P 500 returns over the inter-Greenbook period are moderately positively correlated with both measures of Tonality. Finally, the positive correlation between Tonality Shock and stock returns indicates that innovations to Tonality are related to the news driving the recent stock returns.

IV. Greenbook Tonality as Contributor to Forecast

A. Univariate Predictor of Forecast Errors

Having established a strong connection between the point forecasts of key economic performance measures and our measure of sentiment from the forecast narrative of the same document, our analysis turns to the central question of interest: does that measure of sentiment, Tonality, provide incremental predictive power for these measures of economic performance? Specifically, does Tonality contain information regarding future GDP growth, unemployment, or inflation beyond what is conveyed by the corresponding point forecasts? To gauge the predictive content of Tonality, we begin by estimating univariate regressions that test whether Tonality has the power to predict Fed staff forecast errors. The dependent variable in each regression is the realized forecast error, while the explanatory variable is Tonality of the narrative from the corresponding Greenbook. For GDP, the forecast error is measured relative to the third quarterly

[12] It might be of interest to some readers how the LM-Net measure of sentiment (based on Loughran and McDonald (2011) dictionary) correlates with Greenbook forecasts. In brief, we find that it has a very similar correlation with the outlook for GDP growth and unemployment. In contrast to Tonality, however, it has a strong positive correlation with inflation and with the current-quarter unemployment.

GDP estimate ("first final") published by the BEA. For CPI and unemployment, we use the initial monthly release values, compiled into the quarterly values. More details are provided in the Internet appendix.

For each economic forecast variable and horizon, OLS is used to estimate the conditional expected forecast error as a linear function of either Tonality or Trend Tonality. In addition, for each forecast, we estimate quantile regressions to provide estimates of the 10^{th} and 90^{th} quantiles of the forecast error distribution, conditional on Tonality. This allow us to gauge whether Tonality signals information about the tails of the forecast error distributions: do downside or upside tail risks to the forecast vary notably with Tonality? Table 2 provides the results from all these regressions, with statistically significant coefficient estimates highlighted in blue.

Focusing first on the OLS regressions in the block of three columns to the left, the first four rows show regressions where the dependent variable is either the one-quarter or four-quarter forecast error for GDP growth. As shown in rows one and three, the OLS coefficients on Tonality are positive and significant for predicting GDP growth forecast errors at both the one-quarter and four-quarter horizons. This implies that, when the text sentiment is more positive, GDP growth will tend to exceed point forecasts, though modest the R-squared statistics of 0.03 and 0.04, respectively, suggest the information content is not very substantial, on average, at either horizons. The second and fourth rows show results when GDP forecast errors are regressed on Trend Tonality. The coefficients are again positive and somewhat larger, reflecting the lower variance of Trend Tonality, but here the explanatory power is more substantial, particularly for the four-quarter forecast, with an R-squared equal to 0.12.

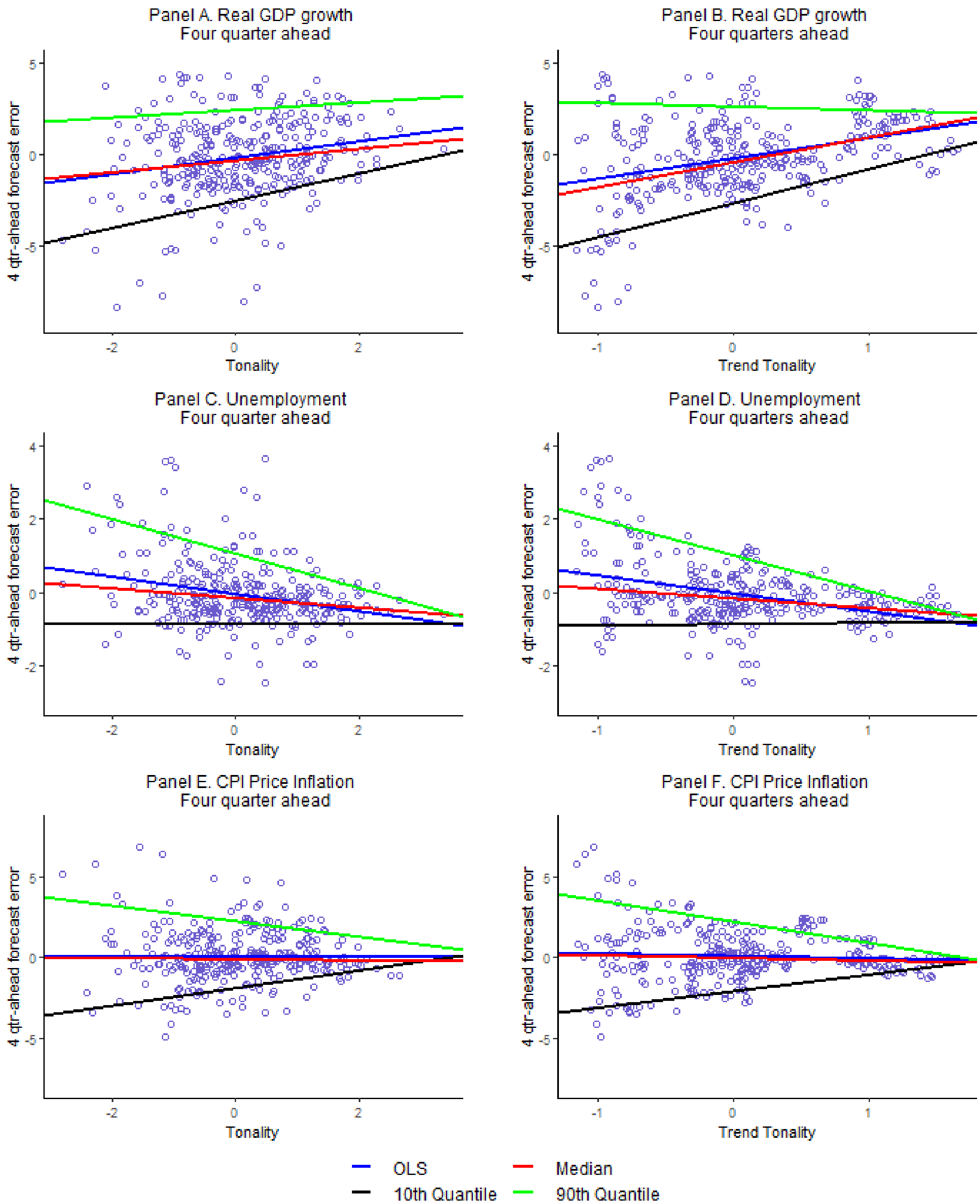
Turning to the remaining OLS regressions, those used to predict unemployment forecast errors are shown in the middle left block of rows. The results here are entirely analogous to the results for GDP, but with coefficient signs reversed. In these regressions, a positive error, (higher than expected unemployment) represents bad news, so results indicate that higher sentiment predicts less unemployment than the point forecast. As in the GDP regression, Trend Tonality again tends to provide a stronger signal of the likely forecast error that does raw Tonality, particularly at the four-quarter horizon, where the R-squared is 0.14. The third set of OLS regression results shown in the bottom left block predict forecast errors for CPI growth. Here we find that Tonality, by either measure, has no predictive power for mean forecast errors.

Regression testing whether Tonality conveys information about the tails of the forecast errors distributions are exhibited in the remainder of the table. The middle three columns show statistics from predicting the 10^{th} quantile of forecast errors, while the rightmost three columns show estimates for the 90^{th} quantile. Considering first the results for GDP growth forecasts, the estimated coefficients from the 10^{th} quantile regression generally are about twice the size of the respective OLS estimates of mean effects, and they are significant at the 1 percent level for both horizons, with R-squared statistics that range as high as 17 percent. As in the OLS regressions, Trend Tonality appears to convey substantially more signal than raw Tonality. In contrast, statistics in the top block of the last three columns (top block) indicate that tonality has no explanatory power for GDP forecast errors in the 90^{th} quantile regressions.[13]

The top two panels of **Figure VIII** show a scatter plot of forecast errors for GDP growth four quarters ahead, respectively, plotted against Tonality to the left and Trend Tonality to the right. The blue line in each figure depicts the OLS regression line while the other lines show the estimated 10^{th}, median, and 90^{th} quantiles of the forecast errors conditioned on its respective version of Tonality. While close inspection reveals the OLS lines to be upward sloping for both measures of sentiment, the more striking pattern is the asymmetry in the signal conveyed, particularly by Trend Tonality: the 10^{th} quantile line is more steeply upward sloping than the OLS line, whereas the 90^{th} quantile line appears to be flat. Consistent with the statistical results in Table 2, this picture shows that lower sentiment predicts realizations that tend to fall short of forecast, but the strongest signal from Tonality is for downside tail risk, gauged by the slope coefficient from the 10^{th} quantile of forecast errors.

[13] An anonymous referee pointed that the higher information content of Tonality for the 10^{th} percentile could be driven from extreme observations. We find similar pattern with the 25^{th} and 75^{th} percentile, where the Tonality has more information in the 25^{th} percentile, than in the 75^{th} percentile.

Figure VIII: Regressions Predicting Forecast Errors: OLS, 10th and 90th Quantiles



Note: The top two panels show scatter plots of forecast errors for GDP growth four quarters ahead, plotted against Tonality on the left and Trend Tonality on the right. The middle and bottom panels show analogous plots for unemployment and inflation forecast errors. In each panel the blue line depicts the OLS regression line while the other lines show the estimated 10th, median, and 90th quantiles of the forecast errors conditioned on Tonality or Trend Tonality.

As illustrated by the quantile regression statistics for unemployment forecast errors, as well as by the scatter plot and regression lines depicted in the middle two panels of Figure VIII, the findings for unemployment forecasts are quite consistent with the results for GDP forecasts. Tonality has strong predictive power for the 90th quantile of forecast errors, when unemployment is much higher than expected, which, again is the tail that reflects bad news. Meanwhile, Tonality has no predictive power for the 10th quantile of unemployment forecast errors. Also analogous with GDP forecast error predictions, the predictive power in the bad news tail is much stronger for Trend Tonality than for raw Tonality, particularly at the four-quarter horizon, where the pseudo-R-squared is 0.26.[14]

The results for CPI forecast errors are, again, quite different. As noted earlier from the OLS regressions, Tonality (and Trend Tonality) has no predictive power for mean CPI forecast errors. However, as indicated by the bottom panels in Figure VIII, the quantile regressions imply that lower Tonality predicts both greater downside risk as well as greater upside risk to inflation, that is, the predominant effect of lower Tonality is higher forecast error variance. Indeed, the third block of rows in Table 2 confirms that these tail effects are statistically significant. It also suggests that, as in the GDP and unemployment regressions, the information content appears to be concentrated in Trend Tonality, rather than the Tonality Shock.

Broadly speaking, these results in some sense mirror key findings in Adrian, et al (2019) and particularly Adams, et al. (2020). These studies examined how the conditional distributions of the same three economic variables varies with financial conditions, measured by the Federal Reserve Bank of Chicago's National Financial Conditions Index (NFCI). In quantile regressions, they find that downside risks to forecasts of GDP and upside risks to forecasts of unemployment increase substantially with less favorable financial conditions. In addition, they find that the financial conditions index conveys no information for predicting mean inflation forecast errors, but that less favorable financial conditions boosts both upside and downside

[14] An anonymous referee suggested trying to corroborate our interpretation of Tonality as a signal of the balance of perceived risks to growth relative to point forecasts by comparing Tonality with similarly timed density forecasts from the Survey of Professional Forecasters (SPF), studied for instance in Clements (2008). We construct measures of the skew, or the relative distance of the 10th and 90th percentile tails in the SPF density forecast distribution from SPF consensus point forecasts, but we find no relationship with Tonality.

risks to inflation forecasts. Similarly, Hennge (2019) analyzed how the conditional distribution of GDP growth varies with the well-motivated measure of economic uncertainty (MacroUnc) introduced by Jurado, et al. (2015) . She finds that, like the NFCI, MacroUnc contains substantial predictive information regarding downside risks to future GDP growth, but little information about upside risks.

B. What Factors Might be Reflected in Tonality?

The similarity of empirical relationships in those three studies to the relationship we find between forecast narrative sentiment and subsequent errors to the staff's economic forecasts suggests that Tonality may be strongly influenced by financial conditions and uncertainty. Indeed, the plausibility of this hypothesis is supported by the strong correlation reported in Table 1 between each of those two variables and Tonality. This raises the question, then, of whether the conditioning information in Tonality simply reflects the information in MacroUnc or NFCI.

This question, as well as other hypotheses regarding the nature of Tonality's predictive information, are tackled in forecast error regressions reported in Table 3, where we simultaneously condition on Tonality and other variables that might influence sentiment or predict forecast errors. For this analysis we focus on GDP and unemployment forecasts, the variables for which Tonality was found to predict mean forecast errors. We focus on the four-quarter forecasts for those two variables, the horizon for which the explanatory power of Tonality and other conditioning variables is highest. Beginning with the effect of Tonality on four-quarter GDP forecast error, the first row shows (once again) the key univariate regression results, the coefficient on Tonality, its p-value and the regression R-squared from the OLS, 10th, and 90th quantile regressions. The subsequent four rows show the analogous statistics for the Tonality coefficient when we control for a competing candidate predictor. The block of regression statistics below pertains to the same set of regressions but with Trend Tonality.

In the case of GDP forecast errors, Tonality by itself is significant with a positive coefficient in the OLS regression and an R-squared of 4 percent. As before the 10th quantile coefficient is double this magnitude with an R-squared of 7 percent, while the 90th quantile coefficient is insignificant. Controlling for MacroUnc, does not alter the estimated effects of

Tonality on the mean forecast error; however, doing so does eliminate the asymmetry in Tonality's marginal effects, with Tonality's coefficient becoming markedly larger and significant in the 90th quantile regressions, but shrinks some in the 10^{th} quantile regression. Its effect is highly significant in all three regressions. This suggests at least some of the asymmetry in the predictive power of Tonality shares reflects a relationship with uncertainty.

Interestingly, as shown in the subsequent line, controlling instead for NFCI leads to very similar results. All told, when we allow MacroUnc or NFCI to, in effect, control for downside risk, Tonality continues to have marginal predictive power, though now roughly similar across the quantiles. The takeaway is that, despite its sizable correlations with MacroUnc and NFCI, and the commonality of their signals regarding downside risk, the predictive information in Tonality for GDP growth is, at least in part, distinct from that signaled by each of these variables.

The other two control variables we examine—the inter-Greenbook period stock return and the Greenbook forecast revision, considered in the subsequent two lines—can be interpretated as signals of recently received information that the staff might have been slow to incorporate into point forecasts. If the explanatory power of Tonality resulted largely due to the forecast being sluggish to adjust but the narrative being more nimble, then including the forecast revision or the recent stock market return might diminish the marginal explanatory power of Tonality. When controlling for inter-Greenbook stock return, we find that the OLS coefficient on Tonality is somewhat lower and only marginally significant, while the regression R-squared rises. In contrast, including stock return has no effect on Tonality's coefficient in the 10^{th} quantile regression, which remains highly significant. Finally, including the forecast revision has no effect on our results.

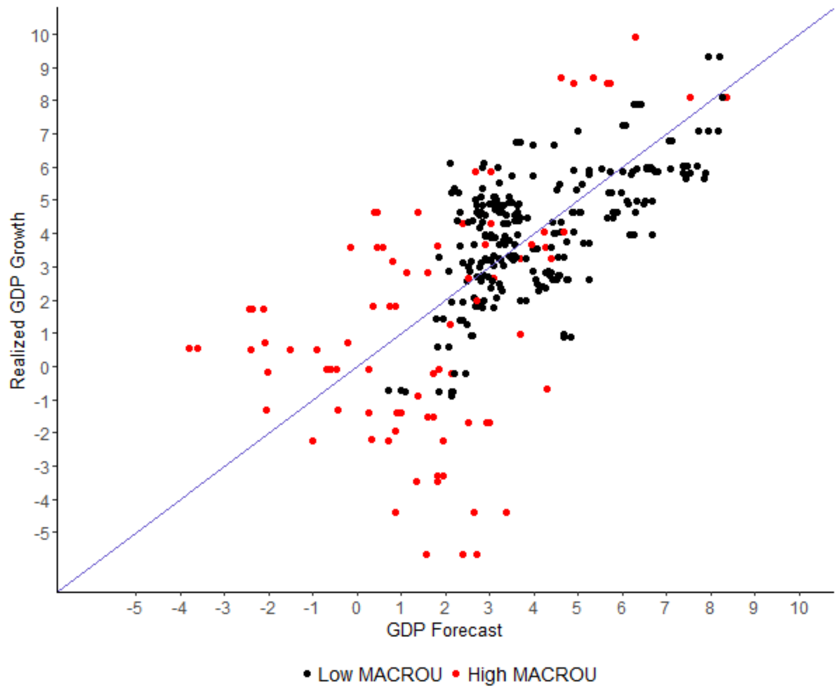
The next block of regressions, which examine the effects of the various controls on our inferences regarding Trend Tonality, finds largely similar results, except that the statistical significance of Trend Tonality always remains. As with Tonality, adding MacroUnc or NFCI appears to boost the estimated marginal effects of Trend Tonality, most notably in the upper quantile regressions. Finally, among all of the control candidates, MacroUnc appears to add the most to the regressions' explanatory power.

For unemployment forecast errors, the results are quite analogous to those in the GDP regressions. The one exception to Tonality's robust marginal effects is that controlling for NFCI in the Unemployment regressions somewhat reduces the magnitude of the negative OLS coefficient on Tonality (or Trend Tonality), leaving it only marginally significant. However, Tonality's effects in the 90th quantile regression remain quite robust after controlling for the NFCI, as are all its OLS and quantile effects when controlling for other variables.

C. When is Tonality most informative?

For forecasts of GDP growth and unemployment, the evidence indicates that Tonality is informative both about the mean expected outcome and, perhaps even more so, about downside risks to economic activity. Now we examine the ex ante conditions under which Tonality is likely to be most informative. In particular, it is useful to consider whether Tonality is more informative about the likely direction of forecast errors when, for instance, uncertainty is relatively high, particularly in light of the incremental predictive information we find when including MacroUnc in Table 3. First, it is useful to gain some perspective on how GDP forecasts and the associated forecast errors are related to macroeconomic uncertainty, as gauged by MacroUnc, which is shown in **Figure IX**. The scatter plot shows realized four-quarter GDP growth (vertical axis) plotted against the associated forecast produced in Greenbook four quarters earlier.

Figure IX: Realized four-quarter GDP Growth versus Forecast



Note: Scatter plot of forecast against realized value of four-quarter GDP growth. Forecasts made when macroeconomic uncertainty (MacroUnc) was in the top quartile of historical values are denoted by red dot. Points that fall on the blue line (with 45 degree slope) when forecasts perfectly align with the realization. Dots far away from the line indicate forecasts with high forecast error.

Observations for which uncertainty is high—specifically, when MacroUnc is within the top quartile of its range—are shown by red dots; the remaining observations, characterized by moderate to low uncertainty, are shown by black squares. The distance of any point from the diagonal line indicates the size of the realized forecast error. A few interesting observations can be drawn from this picture. First, consistent with intuition, forecasts made under high uncertainty tend to result in larger average forecast errors. In particular, the root mean squared forecast error among the red (high uncertainty) observations was 3.4 percent, compared to 1.5 percent among the other observations. Second, it is notable that the vast majority of forecasts projecting subpar growth, such as growth below 2.5 percent, were made when uncertainty was high. What is more, among these observations, the correlation between forecast and realization appears quite low.

We consider the following hypotheses: Does Tonality convey more information when MacroUnc is high? Similarly, does tonality convey more information when the GDP forecast calls for subpar growth, compared to other times? While these hypotheses appear somewhat redundant given our observations from the figure, an inference which is conditioned on the GDP forecast has the attraction that the forecast is plainly observable (to the FOMC committee), whereas MacroUnc is a construct estimated from data that is not all available in real time. These hypotheses are examined by regressing realized four-quarter economic performance, either GDP growth or the change in unemployment, on the respective point forecast and on Tonality or Trend Tonality. This specification is more general than the forecast error regressions in that it allows a parsing of the Greenbook predictive power between the point forecast and the tone of narrative sentiment.

The top section of Table 4 show results for GDP growth. The first three columns show estimates for the low-uncertainty subsample, while the latter three show estimates for the high-uncertainty subsample. For the low-uncertainty sample, the coefficient on the forecast in a univariate regression is 0.71, significantly below the 1.0 hypothesized in standard rationality tests, with a regression R-squared of 0.40. When Tonality is added to the regression, its coefficient is positive and statistically significant, and the R-squared rises to 0.43. Adding Trend Tonality instead, (the 3rd column) provides an even larger increase in the regression explanatory power.

In the high-uncertainty sample, the coefficient on the forecast in the univariate regression is also 0.71, but here the R-squared is only 0.20. Tonality is highly significant when added to the regression (5th column), boosting the R-squared increases considerably. Probably most notable, when Trend Tonality is added to the regression (in column 6), the coefficient on the forecast is no longer significant; at the same time, the coefficient on Trend Tonality is large and highly significant and the regression R-squared jumps to 0.58, even higher than in the low-uncertainty sample. These results strongly indicate that, when uncertainty is high, the sentiment of the narrative is by far a more informative indicator of future economic performance than the forecast itself.[15]

[15] If we instead use the more constrained specification from tables 2 and 3, we similarly draw the inference that Tonality explains substantially more of the forecast error variance in the high uncertainty subsample, but continues to be statistically significant for predicting forecast errors in the low uncertainty subsample.

The lower section of Table 4 shows the analogous results for the unemployment regression. Here, consistent with forecast rationality, the coefficient on the forecasted change in unemployment in univariate regressions is close to unity, in both the low- and high-uncertainty subsamples. Again, however, the predictive content of the forecast as measured by R-squared is more than twice as high in the low uncertainty sample, 0.58 versus 0.25. Moreover, while adding Trend Tonality to the regression in the low-uncertainty sample only marginally improves predictive power, doing so in the high-uncertainty sample doubles predictive power; in contrast to the comparable high-uncertainty GDP forecast regression, the Unemployment forecast does contribute significantly to the regression's predictive power, albeit with a fairly small coefficient.

Table 5 shows the analogous set of regressions, but when the sample split is conditioned on the four-quarter GDP forecast—that is, depending upon whether or not it calls for subpar (below 2.5%) GDP growth. The top section shows regressions predicting GDP growth. Here, the results are similar to, if not more striking than, those in Table 4. In the univariate GDP growth regression, conditioned on the GDP forecast exceeding 2.5%, the coefficient on the GDP forecast is 0.77 with an R-squared of 0.31. In this subsample, adding Tonality or Trend Tonality boosts the R-squared to 0.36 or 0.44, while leaving the coefficient on the forecast near unity and highly significant. In contrast, in the second set of regressions, where the GDP forecast calls for subpar growth, the coefficient estimate on the forecast is only 0.22 and insignificant, while the R-squared is about zero. Adding Trend Tonality, which again has a positive coefficient, boosts the regression R-squared to 0.31.

Finally, as shown in lower panel, we arrive at a similar finding. Here again, when GDP growth is forecast to be subpar, Tonality or Trend Tonality contains substantial predictive power for unemployment, while the unemployment point forecast itself conveys little information. Taken together, these results confirm that the signal from Tonality is indeed most informative when the Greenbook quantitative forecast is calling for subpar growth, and, as indicated by Figure 9, tends to coincide with times of relatively high uncertainty.

D. Do both Negative and Positive words matter?

At this point, we examine a generalization of our sentiment measure to look for potential asymmetry in the effect of sentiment to determine whether the signal in Tonality is similarly driven by variation in negativity or positivity. To do so, we can decompose Tonality into two components, Positivity and Negativity. To construct positivity, we ask how positive is a Greenbook with respect to the average Greenbook by subtracting the average positive score and scale that by the same scaling factor as we scale the score for Tonality (standard deviation of positive minus negative score). This procedure ensures that for each Greenbook the sum of positivity and negativity is equal to Tonality as originally calculated. With Positivity and Negativity we can estimate an unobserved components model on each piece separately. Somewhat surprisingly, while the correlation between changes in these two components is effectively zero, the correlation between their levels is 67 percent. This suggests that, on average during periods with higher usage of words conveying positive sentiment, there also tends to be higher usage of words conveying negative sentiment (relative to neutral words). Here, we test whether both components contribute significantly to the forecast by including them separately in forecasting regressions analogous to those in Table 5.

The results of these tests are shown in Table 6, for the two subsamples and for both GDP and unemployment. For predicting four-quarter GDP growth, the top panel, the coefficient estimates on the two components of Tonality are both statistically significant at the one percent level. We also find the coefficients on Trend Positivity and Trend Negativity to be of very similar magnitude, but oppositely signed, indicating that allowing each piece to enter the separately does little for predictive power. In the unemployment forecast regressions, we again find both pieces of S-Tonality to have significant marginal predictive power. In contrast to the GDP forecast regression, however, here the coefficient on Negativity is materially larger than that on Positivity in both subsamples; thus, separating the two components provides a boost to the adjusted R-squared in each regression.

V. The Relevance of Tonality to the Public

So far, our analysis indicates that the sentiment embedded in the text contains valuable information for Federal Reserve policymakers, over and above that contained in the staff's quantitative forecast. In this section, we investigate whether and how that information, summarized by Tonality, if observed in real time, would have been of value to market participants outside the Fed. In particular, we examine the information content of Tonality along three dimensions. First, does Tonality complement private-sector economic forecasts in a similar fashion? Second, does Tonality help predict monetary policy? Third, does Tonality predict future stock returns? Finally, we take a brief look at whether the sentiment reflected in Greenbook Tonality is signaled to the public in formal FOMC committee public communications.

A. Greenbook Tonality and Blue Chip Forecasts

Does the predictive value of Tonality for Greenbook GDP and unemployment forecast errors reflect some built-in, perhaps conscious, complementarity between the point forecast and the narrative? For instance, does the implied "inefficiency" uniquely apply to Greenbook point forecasts? Alternatively, would Tonality similarly complement publicly available private-sector economic forecasts? This question can be explored using publicly available forecasts produced around the same time as the Greenbook, in particular, by examining whether Greenbook Tonality has similar marginal predictive power conditional on those forecasts.[16]

We use the consensus Blue Chip Financial Forecasts from Wolters Kluwer Legal and Regulatory Solution to conduct this exercise starting in 1980, when this publication begins. To do so, we take the conservative approach of matching up each Greenbook with Blue Chip survey responses published (less than a month) after the Greenbook forecast was produced. This approach guarantees that the Blue Chip forecasters were privy to all the data that was publicly available when the Greenbook narrative was produced.

Table 7 shows the results from regressions analogous to those shown in Table 5, though here Greenbook point forecasts are replaced by the corresponding consensus Blue Chip (BC) forecasts. The top panel shows regressions predicting four-quarter GDP growth conditional on the Blue Chip forecast and Greenbook Tonality, where the sample is divided based on whether the Blue Chip consensus calls for four-quarter GDP growth above or below 2.5%. Despite the

[16] This analysis would seem to bear on the issue of whether the Federal Reserve has more information than the median economic forecaster, as in Romer and Romer (2000) and more recently in Nakamura and Steinsson (2018). However, finding that Greenbook Tonality helps to predict forecast errors in, say Blue Chip forecasts does not necessarily imply that the Federal Reserve has an information advantage, since some Blue Chip forecasters might also produce narratives along with their point forecasts that convey information similar to that in Tonality.

sample being smaller on account of later start date, regression estimates are remarkably similar to those in Table 5. Within the first subsample shown in the first three columns, when expected GDP growth is at least 2.5%, Trend Tonality is statistically significant and complements the BC forecast by boosting the regression R-squared from 0.21 to 0.31. In the below-par growth forecast subsample, the BC forecast has no predictive power (similar to regressions that condition on the Greenbook forecast), but adding Tonality or Trend Tonality boosts the R-squared from 0.01 to 0.16 or 0.33, respectively.

The bottom panel regressions predicting the trajectory of unemployment again produce similar results. These results, together with the those in the top panel, show that the economic signal embedded in Tonality is not materially different when used in conjunction with private-sector forecasts than when used in combination with Greenbook forecasts. Thus, what complementarity, intended or not, between the Fed staff's point forecast and the accompanying narrative would similarly apply to consensus economic forecasts made by outsiders.

B. Tonality as a Predictor of Monetary Policy

Given that Tonality is helpful for predicting economic performance up to four quarters ahead, relative to both Federal Reserve and private sector forecasts, we consider the corollary hypothesis that Tonality has predictive power for monetary policy over a similar horizon. Since higher Tonality tends to signal stronger future economic activity relative to economic point forecasts, a logical consequence one might expect is for higher Tonality to also predict policy interest rates that exceed forecasts. The logic of the hypothesis is as follows: to the extent that Blue Chip consensus forecasts of interest rate policy are connected to Blue Chip consensus forecasts for economic growth through something like a "Taylor rule", then positive economic surprises presaged by Tonality should, in turn, presage positive surprises in the path of policy rates. A key presumption behind this hypothesis is that the effects of such positive economic surprises (or unexpected declines in unemployment rate) are not counterbalanced by downward surprises to inflation, an historically uncontroversial presumption prior to the 2008 financial crises.[17]

[17] The logic for such a connection between the Greenbook forecasts of the federal funds rate and Greenbook forecasts for unemployment would similarly hold; however, the federal funds "forecast" in the Greenbook has not always been chosen to minimize forecast errors. For instance Reifschneider and Tulip (2017) report that the

We test the hypothesis that Trend Tonality helps to predict monetary policy in Table 8. In particular, we regress realized errors in the Blue Chip consensus forecast of the quarterly average Fed Funds rate on value of Trend Tonality at the time of forecast. The first three columns show results for the funds rate forecast at the one quarter, two quarter, and four quarter horizons, respectively. As hypothesized, the coefficient on Tonality is positive and statistically significant, at all three horizons; this indicates that higher (lower) Tonality presages policy rates that tend to exceed (fall short of) Blue Chip forecasts.[18]

The last three columns add a term spread to the regression, specifically, the difference between the nominal one-year Treasury yield and the federal funds rate at the time of forecast. This can be interpreted as a gauge of market expectations for the short-term interest rate trajectory, though imperfectly measured to the extent there are fluctuations in the one-year term premiums. At all three horizons, the coefficient on the term spread is positive and highly significant; moreover, adding it to the regression lowers the still significant coefficient on Trend Tonality at each horizon by as much as half. Thus, it would appear that at least some, though far from all, of the signal for policy rates embedded in Tonality was anticipated by the market and reflected in the slope of the term structure.

C. Tonality as a Predictor of Stock Returns

The evidence presented so far indicates that our measure of sentiment in the Greenbook narrative contains information expected future economic performance that was not incorporated in economists' point forecasts. In addition, it appears to help predict errors in consensus forecasts for monetary policy (the fed funds rate) that are directionally consistent with the expected forecast errors for economic activity. These results beg the question: does Tonality contain information not yet reflected in asset market prices? In particular, can Greenbook Tonality also help predict stock market returns? In what follows, we test whether Tonality has

---

Greenbook traditionally has taken a more "neutral" approach to the Fed funds rate forecast, that it has tended to "condition on [funds rate] paths that modestly rose or fell over time in a manner that signaled the staff's assessment … [of the required] adjustment in policy." This could result in errors in the funds rate forecast being predictable even when forecast errors in economic performance were not. We therefore consider a test of Tonality's predictive power for Blue Chip consensus funds rate forecast errors to have a cleaner interpretation.

[18] On the other hand, the intercept in each case is negative, and the intercept values indicate that the funds rate forecast was on average upward biased by 15 basis points per quarter ahead (assuming S-Tonality was zero on average), indicating that forecasters did not anticipate the downward trend of the funds rate over the sample period.

predictive power for stock returns over periods ranging from 3 to and 12-months, beginning the day after FOMC monetary policy announcements. Our focus on multi-month returns contrasts with most previous studies of news sentiment and stock returns, which mostly document daily or weekly return predictability. One recent exception is Calomiris and Mamayski (2019), which finds that textual information aggregated over a month of news articles can help predict stock returns up to one year ahead.

The precise dating of the periods over which we test for return predictability is determined by FOMC dates; for each return horizon, the return period starts the day after the current-period FOMC policy announcement, and it ends on the day before a future post-meeting FOMC policy announcement. For most of the sample, the prediction period endpoints correspond to the FOMC announcement days following either the 2nd prospective meeting (about three months later), the 4th prospective meeting (about six months later) and the 8th prospective meeting (about one year later). Before 1981, meetings were monthly, so the prediction periods prior to 1981 end on the announcement days following the 3rd, 6th and 12th prospective meetings. For these regressions, the full sample extends back to January 1970.

Table 9 shows coefficient estimates from regressions predicting 3-month, 6-month, and 12-month returns on the S&P 500 composite, each in excess of the yield on the maturity-matched Treasury bill. Shown below each specification are the in-sample adjusted R-squared and an out-of-sample R-squared, the latter simulated starting June 1975 with 64 observations reserved to estimate the initial historical relationship. The baseline regressions in the first three columns condition only on Trend Tonality. As shown, for all three horizons, the coefficient on Trend Tonality is positive and statistically significant. Its magnitude for the 6-month horizon is about double that for the 3-month horizon and for 12-month returns is nearly double again. An increase in Trend Tonality of unity—which amounts to roughly 1.5 standard deviations—predicts a 2.7, 5.1, and 9.0 percent higher return over the three horizons, undoubtedly substantial effects.

The adjusted R-squared statistics for the 3-month, 6-month, and 12-month horizons, are 5, 8, and 12 percent, respectively, which are also quite sizable compared with most stock return predictive regressions in the literature, as summarized, for example, by Welch and Goyal (2008). The out-of-sample R^2 statistics are also positive and quite substantial, in notable contrast with many out-of-sample predictive regressions. If a risk-averse investor were able to take advantage of such information in real time, the gain would be economically meaningful.[19]

Given the positive coefficient on Tonality, a natural interpretation for Tonality's predictive value is that it contains information not fully reflected in stock prices at the time Greenbook is produced but instead is revealed to investors over subsequent quarters. In particular, news of a stronger economy that higher Tonality predicates would presumably be accompanied by news of stronger corporate cash flows as well as a decline in risk premiums. On the contrary, it seems improbable that Tonality's effect would reflect a positive risk premium factor; that would have the odd implication that investors demand a lower risk premium when Greenbook sentiment is more negative. On the other hand, the interpretation that Tonality embeds information that predicts stock prices does not necessarily imply a failure of market efficiency, given that this measure of sentiment is not publicly observable. (Indeed, even Fed staff could not literally observe Tonality and might not have been fully cognizant of the sentiment it reflected.)

While Tonality seems an improbable proxy for the equity risk premium, it could well be correlated with the risk premium.[20] Indeed, Table 1 shows that Tonality is highly correlated with current economic conditions and with forecasts of future conditions. Thus, it could be important to control for economic factors or conditions that are potentially priced but also correlated with Tonality. Specifically, we add controls for the current-quarter projected unemployment rate as well as the forecasted change in the unemployment rate over the next two quarters. Current unemployment is negatively correlated with S-Tonality and also seems likely to be a reasonable proxy for business-cycle-driven variation in the equity risk premium, since risk aversion or

---

[19] Using the evaluation framework of Campbell and Thompson (2007) for a risk-averse investor suggests this would boost expected 6-month returns by 9.1 percent. In particular, the risky asset return can be expressed as the sum of unconditional expected return on the risky asset (μ), the signal (T_t), and a random shock (e) with mean zero and variance σ_e^2. Letting $S = (\mu - r_f)/ ((\sigma_T^2 + \sigma_e^2))^{1/2}$ represent the Sharpe ratio of the risky asset when no signal is observed, and γ represent relative risk-aversion, then the gain in expected return from observing the signal is equal to $\frac{R^2}{(1-R^2)} \frac{(1+S^2)}{\gamma}$. Using 0.26 as the 6 month Sharpe ratio (S), consistent with the Sharpe ratio on stocks over the 1927-2009 period, we calculate a gain in the expected 6-month return of 9.6 percent.

[20] The results for return predictability are just as strong, if not stronger, if we control for dividend yield or other standard predictors.

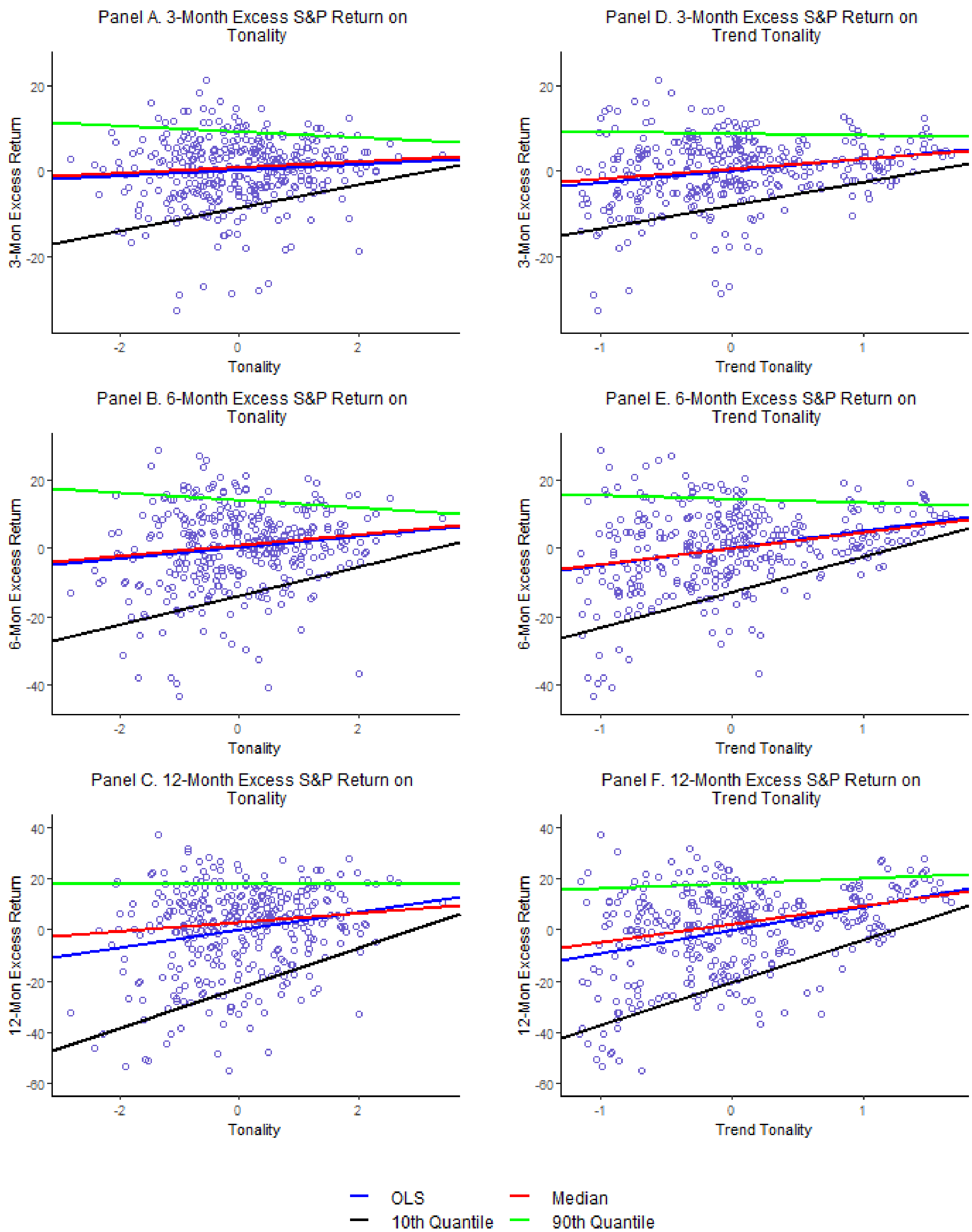
perceived risk are arguably linked to employment prospects.[21] Indeed, a perceived-risk interpretation is invoked by Schmidt (2016) as the rationale behind the return predictability he documents for initial unemployment claims (which are strongly correlated with the Greenbook forecast of Current Unemployment). Another influence on Tonality we control for is the outlook as reflected in the Greenbook point forecast. For this purpose, we use the projected two-quarter change in unemployment.[22]

As shown in columns 4, 5, and 6 of Table 9, when excess returns are regressed on Trend Tonality and these other two factors, we find Trend Tonality to have even greater predictive power for stock returns over all three horizons, with larger and more highly significant coefficients than in the first three columns. At the same time, current unemployment has a positive and significant coefficient, consistent with a risk factor interpretation. The forecasted 2-quarter change in unemployment is also significant with a positive coefficient, consistent with the presumption that the risk premium is higher when the outlook is expected to deteriorate. With the added controls, the in-sample adjusted R-squared statistics are roughly double and out-of-sample R-squares by an even larger amount compared to those from the analogous regressions

Figure X: Regressions Predicting S&P 500 Returns: OLS, 10th and 90th Quantiles

[21] This point is made in the seminal paper on the equity premium by Constantinides and Duffie (1996).

[22] Results are very similar if we use the 2-quarter GDP forecast. For this exercise, we use the two-quarter, rather than the four-quarter, forecast for unemployment because the two-quarter forecast is available going back to 1970, while using the four-quarter forecast would shorten our sample a few years. Since the two are highly correlated, for the shorter sample, it is pretty much immaterial which horizon is used.



Note: Scatter plots of 3-month, 6-month, and 12-month excess stock returns plotted against either Tonality or Trend Tonality from associated Greenbook narrative. The blue line depicts the OLS regression line while other lines show the estimated 10th, median, and 90th quantiles of the forecast errors conditioned on Tonality or Trend Tonality.

with without controls, bolstering the conclusion that the incremental information in Tonality, if available to investors in real time, would have been quite valuable.[23]

Finally, we consider whether the pattern of Tonality's information value for predicting stock returns echoes that for the predictability of economic forecast errors? First, we examine whether the predictive power of Tonality for stock returns is associated with downside risk; that is, is it strongest toward the lower end of potential outcomes—low to negative excess returns—similar to Tonality's signal for the distribution of economic forecast errors? **Figure X** shows quantile regression estimates for 3-month, 6-month and 12-month excess returns, conditioned on Tonality in the left panels and Trend Tonality in the left panels. At all three horizons, we find both measures of Tonality to have their largest predictive effects for returns toward the lower tail of the return distributions, mirroring our findings for macroeconomic predictability.

To examine whether Tonality's predictive power for returns depends upon the forecasted strength of the economy, in Table 10, we estimate the same regressions Table 9 for the 3-month and 12-month horizons, but for subsamples conditioned on the four-quarter point forecast. Comparing results in column 1 to column 3, where excess returns are conditioned only on Trend Tonality, we do find somewhat greater predictive power in the subsample with sub-par expected growth. However, when we also control for current unemployment and the 2-quarter outlook, we find that predictability is much stronger when the forecast calls for subpar growth (column 4) than otherwise (column 2). Taking the 3-month return horizon for instance, when the forecast calls for subpar growth, the coefficient on Trend Tonality is 7.25 and the adjusted R-squared 0.21, compared to a coefficient of 2.48 and an adjusted R-squared of 0.05 in subsample with stronger growth forecasts. For the 12-month horizon, the results are even more dramatic, indeed, almost embarrassingly so.

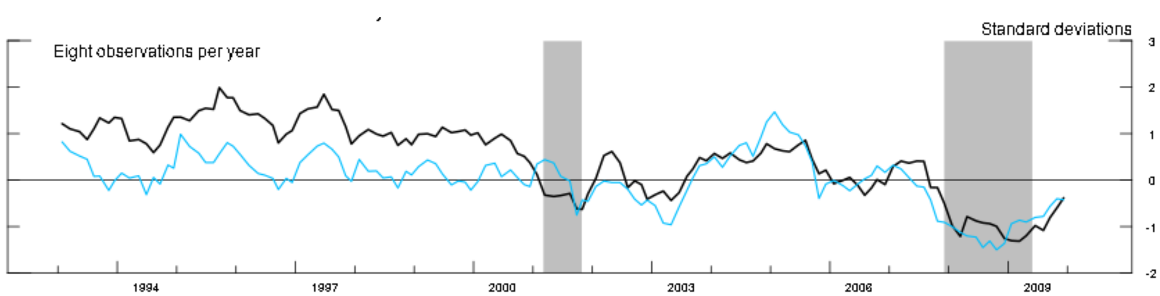
D. Is Greenbook Tonality Communicated to the Public?

[23]The statistical strength of the coefficient on current unemployment rate and the resultant rise in R-squared might seem implausible, as it raises the question as to why others have not documented this to be such a powerful return predictor. Although not shown in the table, when we drop Trend Tonality from the regression, the marginal predictive power of current unemployment drops markedly, and the out-of-sample R-squared statistics turn negative, indicating that the power of this variable comes through only when controlling for Tonality.

In February 1993, the committee began issuing minutes of its deliberations after a delay of several weeks but prior to the subsequent meeting. In February 1994, the FOMC committee began releasing relatively terse statements explaining its actions or stance, at first sporadically and then after every meeting starting May 1999. Stekler and Symmington (2016) and Ericsson (2016) gauge the degree of optimism in FOMC meeting minutes by manually scoring word usage and show that the resulting gauge of sentiment is highly correlated with near-term GDP growth forecasts in the corresponding Greenbooks. In a similar vein, we consider whether the sentiment as measured by Tonality of the Greenbook narrative is signaled through FOMC communications by measuring the Tonality of the two regular public communications: (i) FOMC statements and (ii) minutes of the FOMC meetings.

For each set of communications, Tonality is measured by counting positive and negative word usage in those documents and normalizing using the tdf-if routine, as in our analysis of the Greenbooks. The resultant time series for the Tonality of the FOMC statements is uncorrelated with Greenbook Tonality (0.04 for full sample, same as the post-May 1999 sample). In contrast, the correlation of 0.51 between Minutes Tonality and Greenbook Tonality, depicted visually by a plot of the two time series in Figure XI, indicates quite substantial commonality. Moreover, analogous UCSV estimates of Trend Minutes Tonality is even more tightly connected with Trend Tonality, with a correlation of 0.79.

Figure XI: Minutes versus Greenbook Trend S-Tonality



Note: Shaded regions represent NBER-dated recessions. The black line is the Greenbook Trend S-Tonality. The same smoothing parameters are applied to the minutes' Tonality, shown by the blue line. The minutes are matched to the corresponding Greenbook for this plot.

While a more detailed analysis of the relationship between Greenbook and Minutes Tonality is beyond the scope of this study, this figure provides fairly strong evidence to suggest that the FOMC committee both internalizes and communicates to the public a good deal of the sentiment conveyed in the Greenbook narrative. In light of this, it should not be surprising that a cursory analysis (not shown here) indicates that, over the subsample during which Minutes Tonality is available, a good deal of the predictive power of Greenbook Tonality for funds rate policy and for stock returns carries through to Minutes Tonality.

VI. Summary, Interpretation, and Conclusions

The predictive value of Greenbook Tonality for unemployment and GDP growth, even when conditioning on the Greenbook forecast for those variables, suggests that an important element of economic forecasting is embodied in the accompanying narrative. Having shown that Greenbook Tonality also helps to predict forecast errors for the Blue Chip consensus, it seems clear that the information embedded in the text has broader value than simply as a complement to the Greenbook forecast. The analysis also indicates that very little, if any, of the predictive ability of Tonality reflects either stickiness in the forecast or information signaled by recent stock price movements. What is more, the predictive information in Tonality is somewhat distinct from, even if related to, that signaled by measures of either macroeconomic uncertainty or financial conditions. Indeed, the predictive power of the narrative appears to be strongest at times of high uncertainty, which coincide with times of low economic growth expectations.

The finding that Tonality predicts errors in Blue Chip funds rate forecasts indicates that Tonality conveys policy-relevant information. The finding that Tonality predicts future stock returns should not entirely surprising once we have established its ability to predict unexpected economic growth, but the high R-squared for excess returns, both in and out of sample, is quite remarkable. Given that lower Tonality, i.e., negative sentiment, predicts both greater downside economic risks as well as much lower-than-average returns, the time varying return documented here does not seem to reflect expected compensation for perceived risk. Rather, these results suggest that equity prices do not impound all the information about the potential evolution of the economy that is impounded in the forecast narrative.

The evidence presented in this paper argues for including other narrative information that forecasters are relaying along with their quantitative point forecasts when examining forecast effectiveness or how economic agents update their beliefs. Doing so will require preserving or obtaining the narrative accompanying the forecasts. Quantile regressions for forecast errors suggest that the information in that narrative may be focused on the likelihood of negative tail outcomes. While we have shown that the tone of the narrative that accompanies the Fed's economic forecast is informative, our findings raise some questions. Perhaps one of the more intriguing is whether the Federal Reserve's staff forecast narrative is special in this regard, or whether the narrative from other economic forecasters embeds similar information. In a different vein, given that the recent source of uncertainty, the pandemic, is so different from the past, would it not be more problematic to extrapolate signals in the narrative of late based on past relationships? Finally, we readily acknowledge that this study paper uses a relatively coarse measure of textual information. As suggested by other recent research, deeper and more targeted textual analysis could lead to deeper insight into the nature of economic forecasts.

Appendix A: Text analysis

We used the Harvard psycho-social dictionary as the base dictionary, but exclude words that have special meaning in an economic forecasting context, which leaves us with 231 positive and 102 negative words, which are listed below.

List of 231 positive words

assurance	confident	exuberant	joy	prominent	Satisfactory	unlimited
assure	constancy	facilitate	liberal	promise	Satisfy	upbeat
attain	constructive	faith	lucrative	prompt	Sound	upgrade
attractive	cooperate	favor	manageable	proper	Soundness	uplift
auspicious	coordinate	favorable	mediate	prosperity	Spectacular	upside
backing	credible	feasible	mend	rally	Stabilize	upward
befitting	decent	fervor	mindful	readily	Stable	valid
beneficial	definitive	filial	moderation	reassure	Stable	viable
beneficiary	deserve	flatter	onward	receptive	Steadiness	victorious
benefit	desirable	flourish	opportunity	reconcile	Steady	virtuous
benign	discern	fond	optimism	refine	Stimulate	vitality
better	distinction	foster	optimistic	reinstate	Stimulation	warm
bloom	distinguish	friendly	outrun	relaxation	Subscribe	welcome
bolster	durability	gain	outstanding	reliable	Succeed	
boom	eager	generous	overcome	relief	Success	
boost	earnest	genuine	paramount	relieve	Successful	
bountiful	ease	good	particular	remarkable	Suffice	
bright	easy	happy	patience	remarkably	Suit	
buoyant	encourage	heal	patient	repair	Support	
calm	encouragement	healthy	peaceful	rescue	Supportive	
celebrate	endorse	helpful	persuasive	resolve	Surge	
coherent	energetic	hope	pleasant	resolved	Surpass	
comeback	engage	hopeful	please	respectable	Sweeten	
comfort	enhance	hospitable	pleased	respite	Sympathetic	
comfortable	enhancement	imperative	plentiful	restoration	Sympathy	
commend	enjoy	impetus	plenty	restore	Synthesis	
compensate	enrichment	impress	positive	revival	Temperate	
composure	enthusiasm	impressive	potent	revive	Thorough	
concession	enthusiastic	improve	precious	ripe	Tolerant	
concur	envision	improvement	pretty	rosy	tranquil	
conducive	excellent	inspire	progress	salutary	tremendous	
confide	exuberance	irresistible	progressive	sanguine	undoubtedly	

List of 102 negative words

adverse	dim	feeble	mishap	struggle
afflict	disappoint	feverish	negative	suffer
alarming	disappointment	fragile	nervousness	terrorism

apprehension	disaster	gloom	offensive	threat
apprehensive	discomfort	gloomy	painful	tragedy
awkward	discouragement	grim	paltry	tragic
bad	dismal	harsh	pessimistic	trouble
badly	disrupt	havoc	plague	turmoil
bitter	disruption	hit	plight	unattractive
bleak	dissatisfied	horrible	poor	undermine
bug	distort	hurt	recession	undesirable
burdensome	distortion	illegal	sank	uneasiness
corrosive	distress	insecurity	scandal	uneasy
danger	doldrums	insidious	scare	unfavorable
daunting	downbeat	instability	sequester	unforeseen
deadlock	emergency	interfere	sluggish	unprofitable
deficient	erode	jeopardize	slump	unrest
depress	fail	jeopardy	sour	violent
depression	failure	lack	sputter	War
destruction	fake	languish	stagnant	
devastation	falter	loss	standstill	

Appendix B: Data

In this appendix we provide methodology and source for constructing our dataset. For each set of variables – Tonality, Economic (outcome) variables, Federal funds rate variables, Forecast revisions, Monetary Policy announcement variables, Asset prices and Recession indicators we outline our methodology and source data.

1. *Tonality Variables*

All measures of Tonality are built using text of the Greenbook. Prior to the reorganization of the Greenbook in August of 1974, when it was split into two parts, we use the Recent Developments and Outlook for Domestic Economic Activity portion of Greenbook starting in 1970. Thereafter we use Greenbook Part 1 until December 2009. Of this text, we specifically use the Recent Developments and Outlook for Domestic Economic Activity portion.

Tonality is the number of positive and negative words in a text using a tf-idf weighting scheme from the previous 40 Greenbooks normalized to have mean 0 and standard deviation 1.

Positivity and *Negativity* are the normalized number of positive and negative words respectively using the same tf-idf weighting as Tonality.

Trend versions of Tonality variables are the exponentially weighted moving averages (EWMA) of the normalized Tonality variables with the weighting parameter chosen to maximize fit. The trend measure is fitted over two periods divided at the beginning of 1981, when the frequency of observations changes from 12 to 8 times a year. They are then appended together.

Tonality Shock is equal to *Tonality* variable – *Trend* variable.

2. *Economic Variables*

Historical realized values

The realized values ("actuals") for the economic indicators are real gross domestic product (RGDP), unemployment and inflation as gauged by the consumer price index (CPI) are drawn from the Philadelphia Fed's real-time data set (Croushore and Stark 2001). For GDP, we use the third monthly estimate ("first final") published by the BEA. For CPI and unemployment we use the initial monthly release values, compiled into the quarterly values. We transform the real time data vintages as RGDP growth, CPI growth, and change in unemployment rate. Fed staff forecasted GNP instead of GDP till 1990 and GNP deflator instead of CPI until 1980, hence we use GNP growth and GNP deflator growth accordingly.

The base value for the GDP growth rate is the GDP from the previous quarter at the time of the publication of the Greenbook. Act_RGDP_{-1} is the value of *RGDP* from the previous quarter and $RGDP_i$ is the value of *RGDP i* quarters into the future. We then compute the *i* quarters ahead cumulative GDP growth as following:

$Act\_RGDP\_growth_i = 100 * ((RGDP_i / RGDP_{-1}) - 1)$

Similarly, the unemployment change, we use the quarter prior to the Greenbook publication as base value. $Act\_Unemployment_{-1}$ is the value of *Unemployment* from the previous quarter and $Unemployment_i$ is the value of *Unemployment i* quarters into the future. We then compute the *i* quarters ahead unemployment change as following:

$Act\_Unemployment\_change_i = \text{Unemployment}_i - \text{Unemployment}_{-1}$

Growth in CPI is instead calculated using the contemporaneous CPI. Act_CPI_0 is the value of *CPI* from the current quarter and CPI_i is the value of *CPI i* quarters into the future. We then compute the *i* quarters ahead cumulative GPI growth as following:

$Act\_CPI\_growth_i = 100 * ((Act\_CPI_i / Act\_\text{CPI}_0) - 1)$

Staff Forecasts

All data for staff forecasts of RGDP, unemployment and CPI are from the Greenbook forecast dataset published by Federal Reserve Bank of Philadelphia. We use the forecasts for the previous quarter through four quarters ahead. Forecasts are aligned by the quarter to which the Greenbook is released. With the exception of unemployment rate, data is reported as annualized quarter over quarter percent growth, which we convert to quarterly growth before calculating cumulative growth rates.

$Staff\_RDGP_0$ is the staff's projection of the growth from the previous quarter to the current quarter of RGDP. $Staff\_RGDP_i$ is equal to the projected Q/Q growth *i* quarters into the future. We then compute the *i* quarters ahead cumulative GDP growth as following:

$Staff\_RGDP\_growth_i = \prod_{k=0}^{i} Staff\_RGDP_k$

$Staff\_Unemployment_{-1}$ is the staff's projection for the unemployment rate in the previous quarter and $Staff\_Unemployment_i$ is equal to the staff's projection for the unemployment rate *i* quarters ahead. We then compute the *i* quarters ahead unemployment change as following:

$Staff\_Unemployment\_change_i = Staff\_Unemployment_i - Staff\_Unemployment_{-1}$

$Staff\_CPI_0$ is the staff's projection for the change in CPI from the previous quarter to the current quarter. $Staff\_CPI_i$ is equal to the projected Q/Q growth *i* quarters into the future. We then compute the *i* quarters ahead cumulative CPI growth as following:

$Staff\_CPI\_growth_i = \prod_{k=1}^{i} Staff\_CPI_k$

Blue Chip Forecasts

The Blue Chip forecasts for RGDP, unemployment and CPI are from the consensus estimates from the Blue Chip Economic Indicators publication from 1992 until 2009. The forecast periods are aligned by the month of the Blue Chip public release. In order to match Blue Chip forecasts to Greenbook release dates, the 15th of the month is used as a cutoff. If the Greenbook release date is on or before the 15th of the month, the Blue Chip forecast will be from the same month. In the other case, the next month's Blue Chip forecast will be used. In the event the next month is also the next quarter, one less forecast period is used in order to preserve a constant forecast quarter. After making this adjustment, Blue Chip growth and change variables are constructed in analogous fashion to the variables for the staff forecast.

$BC\_RGDP\_growth_i = \prod_{k=0}^{i} BC\_RGDP_k$
$BC\_Unemployment\_change_i = BC\_Unemployment_i - BC\_Unemployment_{-1}$
$BC\_CPI\_growth_i = \prod_{k=1}^{i} BC\_CPI_k$

3. Federal Fund Rate Variables

Actuals

Until December 16th 2008, we use the target Fed funds rate. Thereafter we use the midpoint of the upper and lower range of the target Federal funds rate. Since the forecasts predict the average rate, we use the average target rate over the entire quarter.

$Act\_FedFunds_{-1}$ is equal to the average Fed funds rate in the previous quarter. $Act\_FedFunds_i$ is the average rate *i* quarters into the future. We define the change in Fed funds rate as follows:

$Act\_FedFunds\_change_i = Act\_FedFunds_i - Act\_FedFunds_{-1}$

Blue Chip Forecast

Blue Chip projections for the Fed funds rate are the consensus estimates from the Blue Chip Financial Forecasts publication from 1992 until 2009. As with economic indicator variables, the Blue Chip forecast is matched to the current Greenbook based on whether or not the Greenbook release date was on or before the 15th of the month. We define the Blue Chip Fed funds variables in the same manner as the staff variables.

$BC\_FedFunds\_change_i = BC\_FedFunds_i - BC\_FedFunds_{-1}$

4. Revisions

We create revision variables for both the Staff and Blue Chip forecasts. Revisions are defined as the difference between the current forecast and the previous forecast for the same period. In the case that the Greenbook release date is in the first month of the quarter, the forecast from the period before will use one additional forecast period in order to maintain the quarterly alignment. For example, in January the revision for a 1-quarter ahead forecast will be calculated as the current 1-quarter ahead forecast minus the

December meeting's 2-quarter ahead forecast. We define the revision for the *i* quarter ahead projection at meeting *t* as follows:

$Revision_{t,i} = Forecast_{t,i} – Forecast_{t-1,i}$

5. Asset Price Variables

We calculate return as the excess of the CRSP S&P 500 return index from the maturity-matched Treasury bill. We also calculate the return from the closing price on day of current meeting to 2, 4 and 6 meetings ahead, roughly corresponding to 3, 6, and 12 months ahead respectively. Stock returns are downloaded from Wharton Research Data Services and are provided by Center for Research in Security Prices, CRSP 1925 US Indices Database, Wharton Research Data Services, http://www.whartonwrds.com/datasets/crsp/.

$SPret_{i,j}$ is equal to the return of the S&P 500 from the *ith* to the *jth* FOMC Date.

$Current\ Unemployment_i$ is the Staff's projection for the current unemployment rate.

Dividend Yield is the 12-month dividend divided by the S&P 500 index value of the previous month (available from Welch and Goyal (2008) and its update).

References

Adams, Partick, Tobais Adrian, Nina Boyarchenko, and Domenico Giannone. 2020. *Forecasting Macroeconomic Risks.* Discussion Paper, London: Centre for Economic Policy Research .

Adrian, Tobias, Nina Boyarchenko, and Domenico Giannone. 2019. "Vulnerable growth." *American Economic Review* 109 (4): 1263-89.

Aggarwal, Raj, Sunil Mohanty, and Frank Song. 1995. "Are Survey Forecasts of Macroeconomic Variables Rational?" *The Journal of Business* 68 (1): 99-119.

Andrade, Philippe, and and Hervé Le Bihan. 2013. "Inattentive professional forecasters." *Journal of Monetary Economics* 60 (8): 967-982.

Asquith, Paul, Michael B. Mikhail, and Andrea S. Au. 2005. "Information content of equity analyst reports." *Journal of Financial Economics* 75 (2): 245-282.

Atkeson, Andrew, and Lee E. Ohanian. 2001. "Are Phillips curves useful for forecasting inflation?" *Federal Reserve Bank of Minneapolis. Quarterly Review* 2-11.

Bai, Jushan, and Pierre Perron. 2003. "Computation and analysis of multiple structural change models." *Journal of Applied Economterics* 18 (1): 1-22.

Baker, Scott R., Nicholas Bloom, and Steven J. Davis. 2016. "Measuring economic policy uncertainty." *The Quarterly Journal of Economics* 131 (4): 1593-1636.

Calomiris, Charles W., and Harry Mamaysky. 2019. "How news and its context drive risk and returns around the world." *Journal of Financial Economics* 133 (2): 299-336.

Campbell, John Y. 1987. "Stock returns and the term structure." *Journal of Financial Economics* 18 (2): 373-399.

Campbell, John Y., and Samuel B. Thompson. 2007. "Predicting excess stock returns out of sample: Can anything beat the historical average?" *The Review of Financial Studies* (21): 1509-1531.

Carvalho, Carlos, Eric Hsu, and Fernanda Nechio. 2016. "Measuring the effect of the zero lower bound on monetary policy." *Federal Reserve Bank of San Francisco Working Paper* 1-32.

Clark, Todd E., Michael W. McCracken, and Elmar and Mertens. 2020. "Modeling time-varying uncertainty of multiple-horizon forecast errors." *Review of Economics and Statistics* 102 (1): 17-33.

Clements, Michael P. 2008. "Consensus and uncertainty: Using forecast probabilities of output declines." *International Journal of Forecasting* 24 (1): 76-86.

Clements, Michael P., and J. James Reade. 2020. "Forecasting and forecast narratives: The bank of england inflation reports." *International Journal of Forecasting* 36: 1488-1500.

Coibion, Olivier, and Yuriy and Gorodnichenko. 2015. "Information rigidity and the expectations formation process: A simple framework and new facts." *American Economic Review* 105 (8): 2644-78.

Coibion, Olivier, and Yuriy Gorodnichenko. 2012. "What can survey forecasts tell us about information rigidities?" *Journal of Political Economy* 120 (1): 116-159.

Constantinides, George M., and Darrell Duffie. 1996. "Asset pricing with heterogeneous consumers." *Journal of Political Economy* 104 (2): 219-240.

Croushore, Dean, and Tom Stark. 2001. "A real-time data set for macroeconomists." *Journal of Econometrics* 105 (1): 111-130.

Da, Zhi, Joseph Engelberg, and Pengjie Gao. 2014. "The sum of all FEARS." *The Review of Financial Studies* 28 (1): 1-32.

D'Agostino, Antonello, and Karl Whelan. 2008. "Federal Reserve Information during the Great Moderation." *Journal of the European Economic Association* 6 (2-3): 609-620.

Das, Sanjiv R, and Mike Y Chen. 2007. "Yahoo! for Amazon: Sentiment extraction from small talk on the web." *Management Science* 53 (9): 1375--1388.

Dovern, Jonas, Ulrich Fritsche, Prakash Loungani, and Natalia and Tamirisa. 2015. "Information rigidities: Comparing average and individual forecasts for a large international panel." *International Journal of Forecasting* 31: 144-154.

Ericsson, Neil. 2016. "Eliciting GDP forecasts from the FOMC’s minutes around the financial crises." *International Journal of Forecasting*, 571-583.

Feng, Xingdong, Xuming He, and Jianhua Hu. 2011. "Wild bootstrap for quantile regression." *Biometrika* 98 (4): 995-999.

Gamber, Edward N, and Julie K Smith. 2009. "Are the Fed's Inflation Forecasts Still Superior to the Private Sector's?" *Journal of Macroeconomics* 31 (2): 240-251.

Garcia, Diego. 2013. "Sentiment during recessions." *Journal of Finance* 68 (3): 1267-1300.

Gilchrist, Simon, and Egon Zakrajšek. 2012. "Credit spreads and business cycle fluctuations." *The American Economic Review* 102 (4): 1692-1720.

Gürkaynak, Refet S., Brian P. Sack, and Eric T. Swanson. 2007. "Market-based measures of monetary policy expectations." *Journal of Business & Economic Statistics* 25 (2): 201-212.

Gürkaynak, Refet S., Brian Sack, and Eric Swanson. 2005. "The sensitivity of long-term interest rates to economic news: evidence and implications for macroeconomic models." *The American Economic Review* 95 (1): 425-436.

Hansen, Stephen, and Michael McMahon. 2016. "Shocking language: Understanding the macroeconomic effects of central bank communication." *International Economic Review* 99: S114-S133.

Hansen, Stephen, Michael McMahon, and and Andrea Prat. 2017. "Transparency and deliberation within the FOMC: a computational linguistics approach." *The Quarterly Journal of Economics* 133 (2): 801-870.

Hengge, Martina. 2019. *Uncertainty as a predictor of economic activity.* Working Paper, Economics Section, The Graduate Institute of International Studies.

Heston, Steven L., and Nitish Ranjan Sinha. 2017. "News vs. Sentiment: Predicting Stock Returns from News Stories." *Financial Analyst Journal* 73: 67-83.

Jones, Jacob, T., Tara, M. Sinclair, and O. Herman Stekler. 2019. "A textual analysis of Bank of England growth forecasts." *International Journal of Forecasting* 1-10.

Jurado, Kyle, Sydney Ludvigson, and Serena Ng. 2015. "Measuring Uncertainty." *American Economic Review* 105 (3): 1177-1216.

Liu, Regina. 1988. "Bootstrap procedures under some non-iid models." *The Annals of Statistics* 16 (4): 1696-1708.

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

Nakamura, Emi, and Jón Steinsson. 2018. "High-frequency identification of monetary non-neutrality: the information effect." *The Quarterly Journal of Economics* 133 (3): 1283-1330.

Newey, Whitney K., and Kenneth D. West. 1994. "Automatic lag selection in covariance matrix estimation." *The Review of Economic Studies* 61 (4): 631-653.

Nordhaus, William D. 1987. "Forecasting Efficiency: Concepts and Applications." *Review of Economics and Statistics* 667-674.

Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, et al. 2011. "Scikit-learn: Machine Learning in Python." *Journal of Machine Learning Research* 2085-2830.

Reifschneider, David, and Peter Tulip. 2017. "Gauging the Uncertainty of the Economic Outlook Using Historical Forecasting Errors: The Federal Reserve's Approach." *Finance and Economics Discussion Series 2017-020. Washington: Board of Governors of the Federal Reserve System* 1-46.

Rogers, John H., and Xu Jiawen. 2019. "How Well Does Economic Uncertainty Forecast Economic Activity?" *FEDS Working Paper* (2019-085). doi:https://doi.org/10.17016/FEDS.2019.085.

Romer, Christina D., and David H. Romer. 2000. "Federal Reserve information and the behavior of interest rates." *The American Economic Review* 90: 429-457.

Rudebusch, Glenn D., and John C. Williams. 2009. "Forecasting recessions: the puzzle of the enduring power of the yield curve." *Journal of Business & Economic Statistics* 27 (4): 492-503.

Schmeling, Maik and Wagner, Christian. 2017. *Does central bank tone move asset prices?* SSRN Working Paper, 1-75.

Schmidt, Lawrence D. W. 2016. *Climbing and Falling Off the Ladder: Asset Pricing Implications of Labor Market Event Risk.* SSRN Working Paper, University of Chicago.

Shiller, Robert, T. 2017. "Narrative Economics." *American Economic Review* 107 (4): 967-1004.

Sinclair, Tara M., Fred Joutz, and Herman O. Stekler. 2010. "Can the Fed predict the state of the economy?" *Economics Letters* 108 (1): 28-32.

Smirnov, Sergey V., and Avdeeva Daria. 2016. "Wishful Bias in Predicting US Recessions: Indirect Evidence." *Higher School of Economics Research Paper.* doi:http://dx.doi.org/10.2139/ssrn.2781923 .

Stekler, Herman, and Hilary Symington. 2016. "Evaluating qualitative forecasts: The FOMC Minutes, 2006-2010." *International Journal of Forecasting* 559-570.

Stock, James H, and Mark Watson. 2007. "Why has US inflation become harder to forecast." *Journal of Money, Credit and Banking* S1: 3-33.

Stock, James H., and Mark W. Waston. 2003. "Forecasting Output and Inflation: The Role of Asset Prices." *Journal of Economic Literature* 788-829.

Stock, James H., and Mark W. Watson. 2007. "Why has US inflation become harder to forecast?" *Journal of Money, Credit and Banking* 39 (s1): 3-33.

Tetlock, Paul C. 2007. "Giving content to investor sentiment: The role of media in the stock market." *The Journal of Finance* 62 (3): 1139-1168.

Welch, Ivo, and Amit Goyal. 2008. "A Comprehensive Look at The Empirical Performance of Equity Premium Prediction." *The Review of Financial Studies* 21 (4): 1455-1508.

Wu, Chien-Fu Jeff. 1986. "Jackknife, bootstrap and other resampling methods in regression analysis." *The Annals of Statistics* 14 (4): 1261-1295.

Zarnowitz, V., and P. A. Braun. 1993. "Twenty-Two Years of the NBER-ASA Quarterly Economic Outlook Surveys: Aspects and Comparisons of Forecasting Performance." In *Business Cycles, Indicators, and Forecasting*, by J., H. Stock and M. W. and Watson. Chicago: University of Chicago Press .

Zarnowitz, Victor. 1985. "Rational Expectations and macroeconomic forecasts." *Journal of Business & Economic Statistics* 3 (4): 293-311.

Zheng Tracy Ke, Bryan T. Kelly and Dacheng Xiu. 2019. *Predicting Returns with Text Data.* Cambridge, MA: NBER Working Paper Series.