Finance and Economics Discussion Series
Divisions of Research & Statistics and Monetary Affairs
Federal Reserve Board, Washington, D.C.

The Power of Narratives in Economic Forecasts

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

2020-001

Please cite this paper as:
Sharpe, Steven A., Nitish R. Sinha, and Christopher A. Hollrah (2020). "The Power of Narratives in Economic Forecasts," Finance and Economics Discussion Series 2020-001. Washington: Board of Governors of the Federal Reserve System, https://doi.org/10.17016/FEDS.2020.001.

The Power of Narratives in Economic Forecasts

Steven A. Sharpe,
Nitish R. Sinha,
and
Christopher A. Hollrah*

First draft: August 30, 2017
Current draft: December 3, 2019

Abstract

We apply textual analysis tools to the narratives that accompany Federal Reserve Board economic forecasts to measure the degree of optimism versus pessimism expressed in those narratives. Text sentiment is strongly correlated with the accompanying economic point forecasts, positively for GDP forecasts and negatively for unemployment and inflation forecasts. Moreover, our sentiment measure predicts errors in FRB and private forecasts for GDP growth and unemployment up to four quarters out. Furthermore, stronger sentiment predicts tighter than expected monetary policy and higher future stock returns. Quantile regressions indicate that most of sentiment's forecasting power arises from signaling downside risks to the economy and stock prices.

JEL codes: C53, E17, E27, E37, E52, G14.
Keywords: Text Analysis, Economic Forecasts, Monetary Policy, Stock Returns

---

* Sharpe (Steve.A.Sharpe@frb.gov) and Sinha (Nitish.R.Sinha@frb.gov) are in the Research and Statistics division at the Federal Reserve Board, 20th Street and Constitution Avenue, NW, Washington DC 20551; Hollrah (Chollrah@umich.edu) is at the University of Michigan. Our views do not necessarily reflect those of the Federal Reserve System or 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 his help in developing the wordlists. We are also thankful to our Board colleagues Jack Bao and Neil Ericsson for detailed discussions. An early version of this paper was circulated as "What's the story: A New Perspective on the Value of Economic Forecasts."

The Power of Narratives in Economic Forecasts

Abstract

We apply textual analysis tools to the narratives that accompany Federal Reserve Board economic forecasts to measure the degree of optimism versus pessimism expressed in those narratives. Text sentiment is strongly correlated with the accompanying economic point forecasts, positively for GDP forecasts and negatively for unemployment and inflation forecasts. Moreover, our sentiment measure predicts errors in FRB and private forecasts for GDP growth and unemployment up to four quarters out. Furthermore, stronger sentiment predicts tighter than expected monetary policy and higher future stock returns. Quantile regressions indicate that most of sentiment's forecasting power arises from signaling downside risks to the economy and stock prices.

I. Introduction

Over the years, many researchers and market participants have questioned the value of macroeconomic forecasts. Nonetheless, substantial resources continue to be devoted to producing detailed economic forecasts. For instance, the Blue Chip Survey of Economic Indicators collects monthly updates of U.S. economic forecasts 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 for 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 these forecasts and presumably their value to financial market participants and policy makers.

In the academic literature, economic 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 in which the quantitative forecasts are embedded, which is often a substantial element 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. It seems quite plausible that policymakers and investors who pay for these forecasts draw significant value from the narratives that accompany individual forecasts.

This study breaks new ground by applying tools from the emerging literature on textual analysis to gauge some of the signal conveyed in 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. In particular, we quantify the degree of optimism versus pessimism embedded in the forecast narrative, which we call the "Tonality" of the text, based upon counts of words that have been classified as positive or negative. The starting point for that classification is the Harvard Psycho-

---

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

social dictionary, which is then fine-tuned by excluding words that have special meaning in an economic forecasting context, such as "demean" and "interest." We find that the resulting measure of forecast narrative sentiment is strongly correlated with accompanying point forecasts for key economic variables, usually with the intuitive sign. In particular, Tonality is positively correlated with forecasts for GDP growth and negatively correlated with forecasted trajectory of the unemployment rate.

The central question we consider is whether our measure of text sentiment has value as a signal of future economic performance. In particular, we test whether this measure of optimism has incremental power, over and above the point forecasts, for predicting key macroeconomic quantities—namely unemployment, GDP growth, and inflation. We hypothesize that positive sentiment helps to predict more favorable economic outcomes, such as higher GDP growth. In short, we find that Tonality has significant predictive power for the change in the unemployment rate and for GDP growth over both a two-quarter and four-quarter forecast horizon. In forecasting regressions, higher text Tonality predicts higher realized cumulative GDP growth, even after controlling for the staff point forecast for GDP growth. Similarly, lower Tonality is also found to presage a higher than expected unemployment rate two quarters and four quarters ahead. In contrast, the directional signal from Tonality to future inflation is ambiguous.

To explore why Tonality contains marginal predictive power for economic performance, we consider two possible hypotheses. First, we test whether this owes to stickiness in the Greenbook point forecasts, that is, forecast revisions that tend to be too conservative. This type of forecast inefficiency was first described by Nordhaus (1987), who points out that "Inefficient forecasts … let the news seep in slowly" and argues that the resultant forecasts errors would be predictable, in part, using recent forecast revisions. More recently, in an analysis of consensus forecasts from the Survey of Profession Forecasters, Coibion, and Gorodnichenko (2015) find evidence of "information rigidity" by showing that forecast revisions for inflation tend to predict future forecast errors in the same direction. They show that such a result can obtain in consensus forecasts even when individual forecast revisions are optimal. Dovern, et al. (2015) finds that revisions of individual forecasts also tend to predict forecast errors in the same direction, though the magnitude of rigidity is smaller than in consensus forecasts.

If such an inefficiency were present in Greenbook point forecasts, this could explain the predictive power of the narrative. In particular, if Greenbook point forecasts were revised somewhat sluggishly, then the text sentiment could be more "nimble" to incorporate new information. This explanation can be tested by adding forecast revisions to prediction regressions. If the predictive power of Tonality owes to sticky point forecasts, then its predictive ability would presumably deteriorate once we control for recent forecast revisions.

Another candidate explanation for Tonality's predictive power is that Greenbook quantitative forecasts are modal rather than mean forecasts, and thus the risks to those forecasts could be systematically unbalanced. For instance, when the perceived likelihood of falling into a recession is higher than average, and thus the mean expected growth rate is substantially lower, a modal forecast for GDP growth might still hew close to the perceived trend growth rate. In this scenario, the text could convey the balance of risks and thus help predict the mean outcome. In other words, sentiment in the text might be particularly informative about tail risks, particularly if quantitative forecasts are modal forecasts. To look for evidence, we estimate quantile regressions on the Greenbook forecast errors and examine whether Tonality is more informative in the upper or lower tail quantiles than around the median forecast error.

In short, we find little evidence that sticky forecasts are the reason for Tonality's predictive power, but fairly strong evidence consistent with the modal-forecast explanation. In quantile regressions, Tonality appears to have its largest estimated effect on GDP forecast errors at the 10^{th} quantile, or near the lower tail of forecast errors. On the other hand, for unemployment forecast errors, the effect of Tonality is largest at the 90^{th} quantile, the upper tail. Together, the results for GDP and unemployment suggests that the marginal predictive value of the text comes disproportionately from its signal of lower-than-forecast real activity, presumably including recessions.[2]

What is more, when we merge our data on Tealbook Tonality together with consensus economic forecasts complied by Blue Chip around the same time, we find that Tonality has very similar power to predict errors in the Blue chip forecasts. And there again, the predictive power

[2] A related candidate for communicating risks to the forecast is a high frequency of words signaling uncertainty, following the seminal work of Baker, Bloom and Davis (2016), but we find that the frequency of "uncertainty" or "uncertain" is generally very low and conveys little information about the likely direction of forecast errors.

appears to be strongest for the unfavorable tail outcomes. This indicates that the information content of Tonality is not simply the consequence of some internal Fed forecasting dynamic, but would have also have value for consumers of private sector forecasts.

In light of the predictive power of Tonality for economic activity (GDP and the unemployment rate), we consider a logical corollary: does Tonality of the text help to predict monetary policy surprises? Consider, for instance, the linkage implied by the Taylor rule. If forecasters consider their Fed Funds forecast to be consistent with their point forecast for the unemployment rate, then upside surprises to the unemployment rate, all else the same, ought to be accompanied by upside surprises to the fed funds forecast. We measure policy surprises as the realized errors in the median Blue Chip forecast of the federal funds rate two and four quarters out, the same horizon that we measure economic forecast errors. We find that Tonality does have significant predictive power for monetary policy; that is, a more optimistic tone in the text presages a higher than anticipated Fed funds rate up to four quarters ahead.

Finally, we ask whether Tonality, if observable in real time, would have conveyed valuable information for stock market investors. Higher tonality predicts stronger future economic outcomes; if that information has not already been anticipated by the market, then we might expect higher Tonality to predict higher stock returns. This hypothesis is complicated, however, by our finding that the news of a stronger economy embedded in Tonality also tends to be accompanied by news of tighter monetary policy, which could temper or even offset any positive stock market effect from the macroeconomic information conveyed by Tonality.

Nonetheless, we find the stock return results to be quite unambiguous, with Tonality having substantial power for predicting positive excess returns on stocks over the 3-, 6- and 12-month holding periods that follow the production of the Greenbook for policymakers. Unlike the conventional interpretation of predictive regressions in the asset pricing literature, our conditioning variable, Tonality, would seem to be a very unlikely proxy for risk or risk aversion, but, rather, a measure of information not yet incorporated in market prices. The positive coefficient on Tonality is consistent with the interpretation that its predictive power arises from its ability to predict cash flow news that investors will receive. That is, higher Tonality predicts subsequent news of a stronger economy, which raises cash flows and presumably lowers investor risk premiums –two factors that would boost stock prices.

A final question we touch upon is whether the sentiment gauged by Greenbook Tonality is transmitted to the public in two subsequent formal FOMC communications, the FOMC statement released following the FOMC meeting and the FOMC meeting minutes released several weeks hence. We find that the Tonality of the relatively terse FOMC statements appear to convey little of that sentiment, whereas Tonality measured from the FOMC minutes correlates fairly strongly with Tonality from the recently-produced Greenbook. Accordingly, FOMC Minutes Tonality appears to have some of the forecasting properties of Greenbook Tonality.

While adding to the literature on the efficacy of economic forecasts, 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. Our paper is similar to a study by Baker, Bloom and Davis (2016) that creates measures of government economic and monetary policy uncertainty by measuring the usage of language in newspaper articles on the subject. It is also similar to a study by Shapiro, Sudhof and Wilson (2017) which finds that sentiment gleaned from the text of newspaper articles outperforms the University of Michigan index of consumer sentiment for predicting macroeconomic series such as output and unemployment, and to Thorsrud (2016) that uses news topics to construct a "nowcast" of the Norwegian economy.

Our study is most closely related to a relatively new area in economics and finance that attempts to quantify narratives, a research agenda recently nudged into the mainstream with the American Economic Association presidential address by Shiller (2017). In particular, our approach is related to recent studies that examine how the tone of newspaper articles helps explain or predict stock market returns beginning with Tetlock (2007), using techniques elaborated upon, for instance, by Heston and Sinha (2017) , Calomiris and Mamaysky (2018) and Ke, Kelly and Xiu (2019). In contrast to these studies, however, our paper measures the narrative written by forecasters rather than the prevailing narrative in popular media. In that sense, our study is related to Asquith, Mikhail and Au (2005), which examines how the sentiment of the text in Wall Street analyst reports explains firms' stock price responses to earnings forecast revisions. Even more similar, Jones, Sinclair and Stekler (2019) quantifies the narrative contained in Bank of England inflation reports and finds that text to contain information that helps predict quarter-ahead inflation.

Also related are recent studies that quantify information conveyed in monetary policy communications and characterize its impacts on markets. Hansen and McMahon (2016) attempt to 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 compare interest rate reactions to FOMC communication before versus during the zero lower bound period. They find that, during the zero lower bound period, positive Fed communications surprises are associated with smaller increases in near-dated government bond yields but similar increases in longer-term yields. Our study differs from these in that we focus on sentiment in the communications between Fed staff and the FOMC committee, information that is only available to the public years later.

Section II describes how we measure Tonality and explores how it co-varies with the point forecasts of key macroeconomic variables in the Greenbook. In section III, we examine the extent to which Tonality conveys information about future macroeconomic conditions not already reflected in point forecasts. In section IV, we explore two potential explanations for why Tonality aids in predicting future economic conditions. Section V examines the relevance of the information in Tonality for market participants, beginning with its ability to predict errors in the Blue Chip consensus forecasts. It then examines Tonality's ability to signal for 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. 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 made public after a 5-year lag. Greenbook forecasts were produced monthly until 1981; thereafter, the 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. For most of our sample, text analysis is based on the text of Greenbook Part 1, the Summary and Outlook, which outlined the forecast. Prior to the document's restructuring in August 1974, we analyze text from the section titled Recent Developments and Outlook for Domestic Economic Activity. Our sample ends in December 2009, the last full year before Greenbook was replaced by Tealbook A, which consolidated Greenbook with some closely related content from the also-retired Bluebook.

We construct an index that quantifies the optimism and pessimism of the Greenbook text, which we refer to as "Tonality." Tonality is equal to the difference between the weighted sum of positive and negative words from our word list. To classify words as "positive" or "negative," we create a custom dictionary of 231 positive words and 102 negative words.[3] To derive our dictionary, we adopt the initial classification of positive and negative words in the widely used Harvard psycho-social dictionary[4] but then exclude words that have a different connotation in the forecasting context. For example, in contrast to the psycho-social dictionary, we do not consider the words "demean" or "hedge" as negative. Positive words in our dictionary include terms like "enthusiasm," "abundant," "enhance," and "successful," whereas examples of negative words include "unrest," "fragile," "trouble," and "gloomy." Our approach is most similar to Tetlock (2007) and Loughran and McDonald (2011), who examine word frequency without trying to gauge the context in which words are used. Like Tetlock (2007), we use the Harvard IV Psychosocial dictionary to classify words; and, like Loughran and McDonald (2011), we use weighted word counts and we cull from the list any words that have domain-specific connotation in economic forecasts.[5]

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,

[3] For the list of positive and negative words, see appendix A.
[4] Tetlock (2007) used Harvard-Psychosocial dictionary to quantify the sentiment in financial news. Da, Engelberg and Gao (2014) use Google searches on select words from this dictionary to quantify fear among U.S. investors.
[5] Using the Loughran-McDonald wordlist instead would yield a very different measure of Tonality, which has only a 24 percent correlation with our measure of Tonality in the Greenbook text, although, separately, positive and negative components of the two measures have 78 percent and 81 percent correlations, respectively.

such as particular economic indicators, or which attempts to identify negations.[6] Such approaches would require a good deal of additional judgment, for instance, on how to classify "nearby" words in text space. It would also necessitate excluding 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.

Figure 1 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.

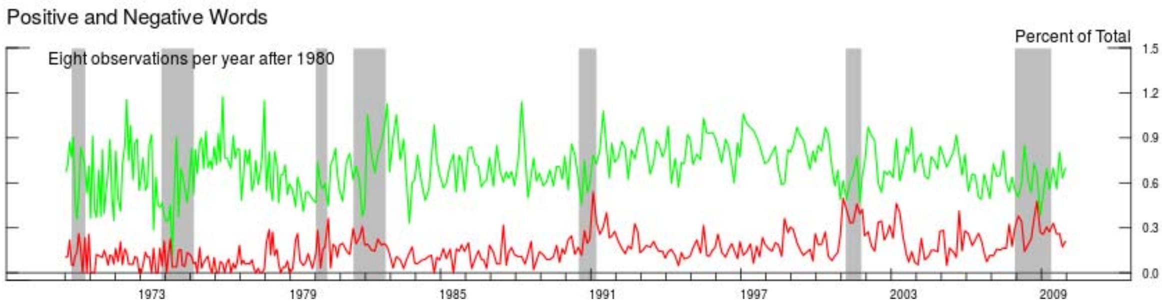
Figure 1: Total words in the Greenbook

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.

Figure 2 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.

[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 2: 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 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 a comparable set of Greenbooks, a weighting scheme commonly known as tf-idf.[7] Specifically, the weight for each word is equal to its current-document frequency (tf) multiplied by the inverse document frequency (idf). For most of our sample, we use the previous 40 Greenbooks as the corpus for obtaining the idf values for a given Greenbook. Early in the sample, for each of the first 40 documents, the corpus is defined to include the first 40 documents.[8]

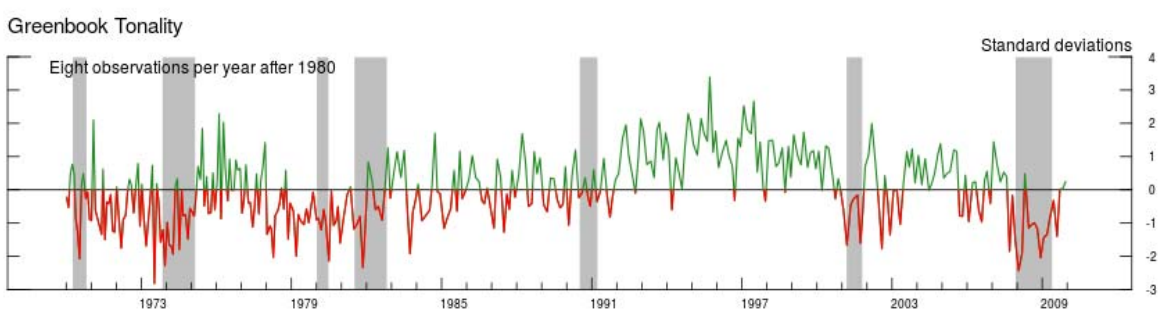
The tf-idf weighing scheme is based on the intuition that infrequently used words are especially informative and so receive relatively high weight in the index, whereas very frequently used words are discounted. Common application of tf-idf scheme would have used the inverse document frequency over *all* 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

[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.
[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.

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 (Pedregosa, et al. 2012) for tf-idf scoring of Greenbooks. Word clouds showing the 50 most prominent positive and negative words in Greenbook during a couple different time periods are shown in the appendix B. Negative words have higher propensity to appear during periods that contain recession.

Figure 3 shows the Tonality index plotted over the full sample period, with positive 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 in negative 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 and 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 much high-frequency movement that is often quickly reversed. To some extent, these fluctuations might reflect noise in our proxy for sentiment.

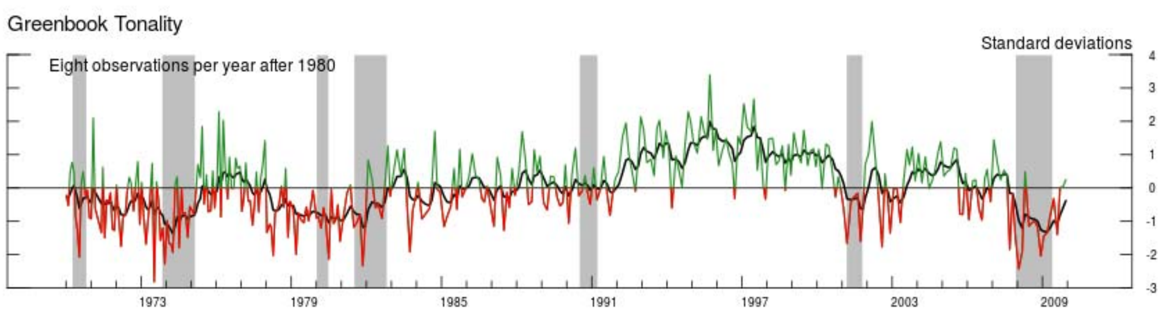
Figure 3: Greenbook Tonality plotted over time



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.

Considering that high-frequency movements could reflect noise, we construct a smoothed measure Tonality which we call "Trend Tonality", as an exponentially weighted moving-average of Tonality. For the post-1980 sample we use a weighting parameter—the decay rate on lagged observations—equal to 0.75; that is, the most recent observation gets a quarter of the weight.[9] For the pre-1981 sample, when Greenbooks were published at a higher frequency (monthly rather than eight per year), we use a somewhat faster decay rate (0.825), calibrated to imply the same calendar-time decay rate. By construction, "Trend" Tonality reflects the slow-moving component of Tonality, while deviations from Trend Tonality reflect possibly temporary shocks. We thus define deviations of Tonality from Trend Tonality as "Tonality Shocks." **Figure 4** shows the resulting times series plot for Trend Tonality, along with (total) Tonality. Not surprisingly, the cyclical pattern in this smoothed measure of sentiment stands out more clearly.

Figure 4: Greenbook Tonality and trend plotted over time



Note: 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 green when negative. Trend Toanlity is the black line overlayed on Tonality and tracks movements in Tonality.

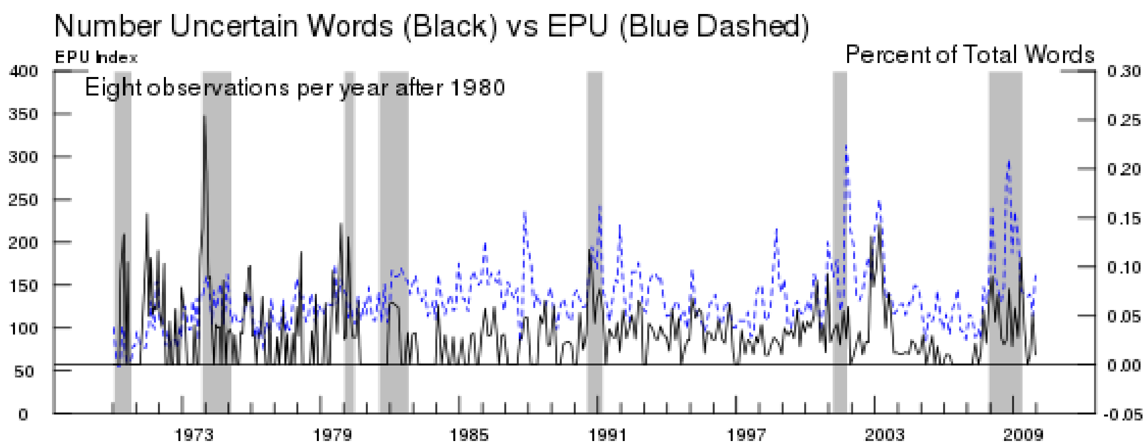
B. Measuring Baker-Bloom-Davis style Uncertainty in Greenbooks

An alternative and increasingly common metric drawn from text analysis is the amount of uncertainty expressed. In their widely cited study Baker, Bloom and Davis (2016) argue that the frequency of "uncertainty" mentions alongside some key words provides a plausible measure of the prevailing uncertainty with respect to economy, monetary policy, or government policy. We follow their methodology but with some tweaks to suit the context of our documents to construct a similarly-styled measure of uncertainty in the Greenbook text. 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

[9] This rate of decay is quite close to the decay rate (of 0.77) that optimizes the one-step-ahead fit between Tonality and Trend Tonality, that is, the decay parameter that minimizes the mean squared distance between the Trend Tonality and the subsequent value of Tonality.

count. The resulting measure is plotted in Figure 5. Notably, early in the sample, there are hardly any mentions of uncertainty; and there are relatively few mentions of uncertainty in the run-up to the 2008 financial crisis.

Figure 5: 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. Relation of Tonality to Concurrent Greenbook Point Forecasts

To examine whether and how text sentiment is related to the associated quantitative forecast, we first examine simple correlations between Tonality and the point forecasts for three key economic performance variables: inflation, the unemployment rate, and GDP growth. The first two constitute the components of the Fed's "dual mandate." The third, GDP growth, is perhaps the most frequently cited summary statistic of economic performance, and its forecasts are presumably closely connected with forecast trajectory of the unemployment rate. For each economic variable we construct a gauge of the two-quarter and four-quarter forecast horizons: in particular, we measure the forecast of cumulative inflation, cumulative GDP growth, and the change in the unemployment rate, each of these over the subsequent two quarters and four quarters out. We also construct the revisions in those forecasts relative to the previous Greenbook. Finally, to gauge the perceived state of the economy at time of forecast, we use the current-quarter forecasts for each metric, that is, the inflation rate, the unemployment rate, and GDP growth.[10]

The correlations of Tonality—both raw Tonality and Trend Tonality—with the current-quarter forecasts, the two-quarter forecasts, and with revisions to two-quarter forecasts are shown in Table 1. Many of the correlations are quite strong, while their signs accord with intuition. Tonality is negatively correlated with measures of inflation and unemployment but positively correlated with measures of GDP growth. What is more, for all three economic variables, the current-quarter and two-quarter forecasts are more strongly correlated with Trend Tonality than with overall Tonality. In contrast, *revisions* to the two-quarter forecasts exhibit similar magnitude correlations with both Tonality and Trend Tonality. This suggests that some of the volatility in Tonality reflects the direction of revisions in the forecast. The final row of Table 1 shows that Tonality is only mildly negatively correlated with the Baker-Bloom-Davis style measure of uncertainty in the text.

Table 2 shows the correlations among the Greenbook forecast variables and with the Uncertainty measure. Perhaps not surprisingly, the correlation of the Unemployment Forecast with the GDP forecast is quite large in magnitude, at -0.86, as are the revisions to these two forecast variables (-0.70). The last row of Table 2 shows that, in general, almost all measures of Greenbook forecast have lower correlation with Uncertainty, as compared to their correlations with Tonality (in Table 1).

We next examine the marginal contributions of the forecast variables for “explaining” Tonality in a multivariate regression context (Table 3). To help keep this preliminary exercise tractable and relatively easy to interpret, we focus only on the (two-quarter) forecasts for the two key components of the Fed’s mandate—inflation and unemployment. We omit the GDP forecast from these regressions because of its very strong negative correlation with the unemployment forecast (shown in Table 2) and the resultant multicollinearity that would introduce.

For the full sample (1972 – 2009), shown in the first column, we find that both the inflation and unemployment forecasts have highly significant marginal explanatory power for

[10] 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, and each have negative coefficients as intuition would predict. Even so, they explain only about 15 percent of the variation of Tonality over the full sample. To determine whether these relationships are structurally stable we use the Bai and Perron (2003) test to look for structural breaks in the multivariate relationship between Tonality and the unemployment and inflation forecasts. As detailed in appendix C, we find strong evidence for a single break, estimated to have occurred in October, 1991.

The second and third columns show the Tonality regression estimates for the early (1972-1991) and late (1991-2009) sub-periods, respectively.[11] The most dramatic disparity between the sub-periods is a change in the sign on Inflation Forecast. Prior to 1992, that forecast has a highly significant negative marginal effect on Tonality, whereas in the later period its coefficient is positive. Although the positive effect of Inflation Forecast post-1991 seems puzzling, it would be consistent with the idea that, after 1991, the Federal Reserve forecast reflected an expectation that inflation would be kept at bay. Perhaps this major structural change in factors behind the sentiment in Fed forecast documents is connected to the so-called "Great Moderation." While researchers commonly date the latter to occur in the mid-1980s, it probably took more time for that change to be fully recognized and reflected in economists' forecasts.

Rounding out the findings from the sample split, we find that the negative coefficient on the unemployment forecast is not statistically significant in the early period but much larger and highly significant in the later period. Moreover the economic forecast variables explain only a small portion of the variation in Tonality over the early period (adjusted R-squared of 7%), but a large share of that variation in the later period (adjusted R-squared of 37%).

The last two columns show the multivariate relationship between the smoothed measure of sentiment, Trend Tonality, and the point forecasts for two forecast variables, using the same sample break as in the previous two columns. Consistent with the conjecture that Trend Tonality is a less noisy measure of sentiment about the outlook, the regression R-squared statistics for both subsamples rise markedly relative to the raw Tonality regressions, to 36% and 59% respectively for the early and late periods. Even so, coefficients are qualitatively similar; in

[11] If we were to incorporate a second break as indicated by the Bai-Perron test, the two later sub-periods (September 1990 to December 2000 and after December 2000) would be qualitatively similar, differing from each other mostly by size of the negative effect of the unemployment rate outlook on Tonality.

particular, the change in the inflation coefficient remains. Perhaps the most notable difference is that the unemployment forecast is a significant determinant of Trend Tonality in both sub-periods.

III. Greenbook Tonality as an Economic Indicator

Having established a strong connection between Tonality and the point forecasts for key economic performance measures in the same document, our analysis turns to a central question of interest: does Tonality have predictive power for such measures of economic performance? For instance, does Tonality contain information regarding future GDP growth that is not fully reflected in the GDP forecast itself? To gauge the predictive content of Tonality, we estimate regressions that test whether Tonality helps to predict the three key economic performance variables we have focused upon. In each regression, the dependent variable is the realized cumulative performance for the variable in question, and the explanatory variables the Greenbook point forecast for the matching horizon as well as Tonality. In light of the structural change in how Tonality of Greenbook text relates to inflation, the inflation forecast regression is estimated on the two separate subsamples, split at October 1991.

The baseline econometric framework for our analysis is adopted from the extensive literature on forecast rationality and efficiency, beginning with studies such as Zarnowitz (1985) and Aggarwal, Mohanty and Song (1995), which examine whether economic forecasts embed systematic errors. The canonical approach involves regressing the realized value of the forecasted variable on the forecast and testing whether the coefficient on the forecast is unity and the intercept is zero. Forecast efficiency tests then examine whether adding other information variables to that regression helps predict the variable of interest.

In our analysis, this suggests the following basic specification:

$$Realized_{t+h} = \alpha + \gamma_h Forecast_{t,t+h} + \beta_h Tonality_t + \varepsilon_{t,h}$$

This represents an efficiency test for the Greenbook forecast because any information reflected in Tonality is presumably observable to the Fed staff producing the point forecast. Note that the specification nests a simple "forecast-error" regression, in which the forecast-error$_{t,t+h}$ (realized less forecast) is regressed on time t Tonality. That specifications would be equivalent to those that follow if we restricted the coefficient on the $Forecast_{t,t+h}$ to unity.

Baseline regressions that examine the predictive content of Tonality for future GDP growth are shown in Table 4. Dependent variables in the first (second) columns for each pair of regression is cumulative realized GDP growth over the subsequent 2 quarters (4 quarters). The first pair of regressions examines the predictive content of the GDP growth forecast by itself. Tonality is added in the second pair of regressions and then is decomposed into Trend Tonality and Tonality Shock in the third pair of regressions. Standard errors are corrected for autocorrelation for $(2*k+1)$ lags for forecast error regressions k quarters out using the automatic bandwidth selection procedure described in Newey and West (1994).

Coefficient estimates on the staff point forecast are 0.96 for 2-quarter GDP growth and 0.80 for 4-quarter growth, neither of which is significantly different from 1.0 at the 5 percent level. The intercept estimates are not statistically different from zero, also consistent with standard tests of rationality. The adjusted R-squared statistics for the two-quarter and four-quarter GDP growth forecasts are 0.55 and 0.40, respectively. When Tonality is added in the second pair of regressions, its estimated coefficient in both cases is positive and significant. Adding tonality boosts the adjusted R-squared only marginally for the 2-quarter forecast, from 0.55 to 0.57, but for the 4-quarter forecast the R-squared rises from 0.40 to 0.45. For the 4-quarter horizon, the coefficient estimate implies that a one-standard deviation increase in Tonality raises expected GDP growth by 61 basis points.

When we split Tonality into its trend and shock components (last two columns), Trend Tonality is the component that contains all of the information for aiding the GDP growth prediction. For both horizons, only the coefficient on Trend Tonality is statistically significant, and it is substantially larger than the Tonality shock coefficient; moreover, the R-squared rises notably again for the four-quarter GDP forecast. To gauge whether users of the Greenbook forecast could have benefited in real time from the information in Tonality, we construct out-of-sample R-squared statistics. For the four-quarter forecast, the improvement in explanatory power from including Tonality is on par with the improvement indicated by the in-sample statistics, which suggests a material real-time benefit. Given indications of a break in the relationship between Tonality and the forecast variables from the preliminary analysis in Table 3, we ran the same regressions on the two subsamples (split at October 1991). As shown in an earlier FEDS Working Paper version of this analysis, we find that coefficient estimates on Tonality in the GDP and Unemployment forecast regressions are similar across subsamples, though Tonality's contribution to forecast performance is much stronger in the latter subsample.

Looking over the full set of regressions, another interesting observation is that the coefficient on the staff forecast declines when Tonality is added to the regression, and even further in the Trend Tonality specification. This suggests that the consumer of these forecast (the FOMC) should have "faded" the Greenbook point forecast somewhat, while putting some weight on the tone of the narrative in Greenbook, as quantified by Tonality. In the traditional research on forecasts, one would conclude that Greenbook GDP point forecasts are not "rational" in the traditional sense that the forecast can be improved upon by incorporating the information that was driving the text sentiment. Of course, at the time, staff forecasters were unable to observe our aggregation of that information into Tonality.

Results from estimating the analogous regressions for the forecasted change in unemployment rate are shown in Table 5. Overall, findings regarding the predictive effects of Tonality are quite similar to those for GDP. Although the improvements in R-squared are modest, Tonality, and particularly Trend Tonality, have predictive power for the change in unemployment over both horizons, with higher Tonality predicting lower unemployment. For instance, an increase in Trend Tonality of 1.0 (about 1.5 standard deviations) predicts a 0.43 point lower unemployment rate.[12]

Owing to the striking contrast in the relationship between Tonality and the inflation forecast across the two subsamples (in Table 3), we estimate inflation forecast regressions for the two subsamples (Panel A and Panel B of Table 6). Tonality and Trend Tonality have negative though only marginally significant coefficients in the early period, but positive and sometimes significant coefficients in the late period. This echoes our findings in Table 3, that is, the sign of the correlation between Tonality and the inflation forecast indicates the direction of the signal

[12] As shown in a previous draft of the working paper, we find that the full sample results masks the much stronger predictive value of Tonality in the post-1991 subsample, especially for Unemployment.

embedded in Tonality for the ultimate realization of inflation. Moreover, in both time periods, that signal in Tonality for future inflation is not particularly robust.[13]

IV. Deeper Dive into Predictive Power of Tonality

In this section we attempt to delve deeper into why Tonality might convey information about future economic performance that is not reflected in the point forecasts. Given the relatively weak and unstable information conveyed by Tonality about future inflation, we focus on Tonality's predictive power for GDP and unemployment. We consider two alternative and somewhat testable hypotheses for *why* Tonality might contain information for these measures of economic performance that is not already reflected in Greenbook point forecasts. First, we consider the hypothesis that point forecasts tend to be sticky, particularly as compared to the accompanying narrative and its tone. According to this hypothesis, the accompanying narrative is not sticky and conveys the information not fully incorporated into the point forecast due to forecast inertia. The second hypothesis we consider is that point forecasts are more akin to modal forecasts than mean forecasts, and that Tonality contains information about the relative importance of upside versus downside risks to the point forecast.

A. Sticky Point Forecasts

To test whether Tonality is informative because point forecasts tend to be sticky, we consider adding to our regressions a variable to serve as a proxy for information available when the forecast is produced, but which might not be incorporated into point forecasts. One such information variable is the revision to the forecast from the previous Greenbook, as first argued by Nordhaus (1987). If Greenbook point forecasts tend to be revised only partway toward their mean-square-error minimizing value, then we would expect that, for instance, adding the revision to the GDP growth forecast to our GDP forecast regression would help predict the forecast error with a positive coefficient. If the text narrative was simply more nimble (less

[13] Though tangential to focus of this paper, it is interesting to note the small and insignificant coefficient estimates on the Staff Forecast in the later period. Indeed, we find that the four-quarter forecast has no predictive power for realized four-quarter inflation. This echoes findings by Atkeson and Ohanian (2001) and Stock and Watson (2007), who show that much less of the variation in inflation has been forecastable since the mid-1980s.

sticky) than the point forecast, then controlling for the forecast revision should reduce the marginal predictive value of Tonality.

Another approach to control for stickiness is to explicitly control for information that has been incorporated into asset prices since the previous forecast. One such measure of recent information about the economy is the recent stock market return (since previous Greenbook), given that stock prices have long been seen as a leading economic indicator (Stock and Waston 2003). If point forecasts are sticky, then tonality might reflect information in stock returns even if it is not fully reflected in the point forecast. Finally, we control for Uncertainty using the Baker-Bloom-Davis style measure constructed from the Greenbook text.

Regressions of realized GDP growth (or unemployment trajectory) on Greenbook point forecasts, Tonality, Staff Revision, and Recent Stock Return are shown in Table 7. Most notably, Recent Stock Return is significant in all specifications, with higher stock return indicating higher realized GDP growth and lower realized unemployment. On the other hand, the coefficient on Staff Revision is sizable only for predicting GDP growth at the 2-quarter horizon, and even there it is not statistically significant. Most importantly, adding the two information variables somewhat reduces the estimated effect and statistical significance of Tonality in the regressions for 2-quarter forecasts. On the other hand, it has little effect on Tonality's predictive power for the 4-quarter horizon, and the estimated effects of Trend Tonality largely hold up for both horizons. Thus, we find weak evidence, at best, for the hypothesis that the predictive power of text sentiment owes to the stickiness of point forecasts.

B. Tonality as an Indicator of Unbalanced Risks to Forecast

Arguably, the typical point forecast in the Greenbook, perhaps in the surveys of professional economic forecasters as well, should be interpreted as representing a modal forecast, rather than a mean forecast that by design minimizes mean squared errors. If so, Tonality could help predict by conveying information about the relative importance of upside or downside risks to a forecast. For instance, it is well-known that quantitative economic forecasts during expansions rarely project recessions. Perhaps Tonality reflects the perceived risk of a recession. One approach to testing whether Tonality's predictive power resides in its ability to signal downside, or upside, risks would be to estimate quantile regressions. In particular, we estimate quantile regressions in which the dependent variable is the realized forecast error in Greenbook—for either the unemployment rate or GDP growth—with Tonality as the key explanatory variable. The first four columns in Table 8a show the key results from quantile regressions where the GDP growth forecast error, for the 2-quarter horizon, is regressed on either Tonality or Trend Tonality. All regressions also control for Recent Stock Return. The remaining columns show results for the 4-quarter horizon.

In each case, the coefficient estimates for the median (50th quantile) regressions are each quite similar to the respective coefficients on Tonality or Trend Tonality in the conventional forecast regressions (Table 4). However, we find that, for both horizons, the coefficient on Tonality or Trend Tonality is larger at the 25th and 10th quantiles. At the 10th quantile in particular, the coefficient on either Tonality measure is about double the coefficient from the median regression. In the case of Trend Tonality for the 2-quarter horizon, the 50th and 10th quantile coefficients are statistically different at the 5 percent significance level. [14] Also worth noting is that Pseudo R^2 statistics generally are higher for the lower quantile regressions, with a maximum value of 13% for the 10th quantile regressions using Trend Tonality. One apparent oddity is that, for the 2-quarter horizon, the Tonality coefficients at the 90th percentile actually have the reverse sign, though only in the case of Trend Tonality is the Psuedo R^2 material. Putting this aside, the quantile regression results imply that Tonality provides a particularly strong signal when GDP growth is going to come in substantially lower than forecast.

Results for quantile regressions on unemployment rate forecast errors (Table 8b) are consistent with GDP forecast error quantile regressions in the sense that the strongest signal from Tonality shows up in the quantiles that up-weight bad economic news. In particular, the largest negative coefficients on Tonality or Trend Tonality are found at the 90th percentile quantile—when Unemployment turns out to be substantially higher than forecast. Here, the difference between the coefficients at the 50th and 90th percentiles is statistically significant in all four specifications. Analogous to the GDP results, the coefficient has the opposite sign at the 10th quantile, the upper end of economic outcomes. Also echoing the GDP results, the explanatory

[14] To obtain the confidence interval for our quantile regression estimate, we follow the smooth block bootstrap procedure developed by Gregory, Lahiri and Nordman (forthcoming) in which we first smooth and taper the variables, choose the block length of 5 periods and bootstrap XY pairs.

power of Tonality is much higher at the 75th and 90th quantiles than elsewhere. Indeed, the Pseudo R^2 for Trend Tonality at the 90th quantile is 19% and 22%, respectively, for the 2-quarter and 4-quarter horizon regressions. Overall, the quantile regressions appear fairly supportive of the conjecture that much of the information conveyed by our measures of text sentiment is related to downside risks to the economy.

V. The Relevance of Tonality to the Public

So far, our analysis indicates that the information embedded in Tonality appears to contain 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 the information reflected in Tonality would be of value to market participants outside the Fed. In particular, we examine the information content of Tonality along four dimensions. First, does Tonality convey similar information relative to private-sector economic forecasts? Second, does Tonality help in predicting monetary policy? Third, does Tonality predict future stock returns? Finally, we take a brief look at whether the sentiment reflected in Greenbook Tonality shows through to formal FOMC committee public communications.

A. Greenbook Tonality and Blue Chip Forecasts

Tonality of the Greenbook narrative has predictive value for GDP growth and unemployment, conditional on the Greenbook forecast; that is, the narrative can predict forecast errors in the point forecast. Does this reflect some built-in, perhaps conscious, complementarity between the point forecast and the narrative? For instance, does this reflect biases in the Greenbook point forecasts induced by some complementary communication built into the forecast narrative? Alternatively, might Tonality have similar information value to the public, whose views are conditioned on publicly available forecasts for GDP and the unemployment rate? This question can be explored using forecast errors constructed from publicly available forecasts produced around the same time as the Greenbook, by testing whether they too can be predicted by Greenbook Tonality.[15]

We use the consensus Blue Chip Financial Forecasts from Wolters Kluwer Legal and Regulatory Solution to conduct this exercise. To do so, we take the conservative approach of matching up the 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 publicly available when the Greenbook narrative was produced. Then we construct Blue Chip consensus forecast errors for 2-quarter-ahead and 4-quarter-ahead forecasts of GDP growth and unemployment, using the Blue Chip forecasts from 1980-2009.

The first four columns of Table 9 show regressions where the dependent variable is the Blue Chip consensus forecast error for 2-quarter and 4-quarter GDP growth. In the first two columns the only regressors are the two components of Greenbook Tonality. In the 3rd and 4th columns, we add the two controls for possible rigidities in forecast adjustment used in testing Greenbook forecast error predictability: the revision to the Blue Chip forecast and the recent stock return.[16] The last four columns repeat the analogous regressions for the Blue Chip unemployment forecast error.

Overall, the results show that Trend Tonality has quite strong predictive power for Blue Chip forecast errors, particularly at the four-quarter horizon. For the GDP forecast errors, the coefficient on Trend Tonality is positive for both forecast horizons, though it is only statistically significant for the longer horizon, where the R-squared is 19%. When the forecast revision and recent stock return are added to these regressions, these two variables have the expected sign (positive) but they are not statistically significant, and their inclusion does not blunt the estimated effects of Trend Tonality.

[15] 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 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.

[16] Like the dependent variables, both of these controls measure changes relative to the time of the Blue Chip forecast published prior to the time of the previous Greenbook, which would be Blue Chip forecasts published either one or two months earlier.

The results for predicting unemployment rate forecast errors are even stronger. Here the coefficient on Trend Tonality is significant at the 1 percent level for both horizons, and the portions of the forecast error variation predicted by Trend Tonality rise to 13 and 25 percent for the 2-quarter and 4-quarter horizons, respectively. When the two controls for forecast rigidities are added we find that both controls are statistically significant (with the expected sign), with the coefficient on revision being quite large in both regressions, in contrast to the results for GDP forecasts. Even so, their inclusion only reduces the coefficients on Trend Tonality by about a quarter. Finally, it is interesting to note that the coefficients on Tonality are of similar size to the estimated coefficients in the Greenbook forecast regressions (tables 4 and 5).

Given those similarities, it is perhaps not entirely surprising that quantile regressions for the Blue Chip 4-quarter forecast errors (Table 10), exhibit a pattern of coefficients quite similar to what we found in quantile regressions for Greenbook forecast errors (Table 8b). For both GDP growth and Unemployment forecast errors, the predictive value of Trend Tonality is strongest around the bad-news end of the forecast error spectrum for both forecast variables. This would seem to reinforce the idea that this pattern of effects is not spurious, and it suggests that Blue Chip forecasts also might be more appropriately considered as modal, rather than mean, forecasts. Moreover, Tonality appears to contain information about the tail of the distribution of possible outcomes.

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 internal Fed forecasts as well as private sector forecasts, we consider the corollary hypothesis that Tonality could have predictive power for monetary policy over a similar horizon. In particular, higher Tonality tends to signal stronger future economic activity relative to economic point forecasts by Fed staff as well as the private sector. As a consequence, one might expect higher Tonality to predicate higher-than-forecast policy rates.

The logic of the hypothesis that Tonality might predict surprises in the Fed funds rate is straightforward, at least for the case of private sector forecasts. 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, presumably a safe assumption so long as the "Phillips curve" is not positively sloped.[17]

The first two columns of Table 11 show estimates from the baseline model for predicting 2-quarter-ahead and 4-quarter-ahead errors in Blue Chip forecasts of the federal funds rate, where the only regressors are two components of Greenbook Tonality. As hypothesized, the coefficient on Tonality is positive and statistically significant at both horizons; thus, higher Tonality presages policy rates that tend to exceed forecast. The 3rd and 4th columns add the two proxies to control for sluggish forecast adjustment, the forecast revision and recent stock return. Here the estimated coefficients on recent stock return are not significant. On the other hand, coefficients on the revision to the funds rate forecast are large and highly significant, echoing the results for the (Blue Chip) forecast error regressions for unemployment.[18] Columns 5 to 8 introduce these controls one at a time and suggest the stock market returns are not as powerful predictors as forecast revisions. Regarding the primary issue at hand, adding the controls does reduce the marginal predictive value of Trend Tonality by about a third, though it remain statistically significant.

We also estimated quantile regressions, not shown, to examine whether the predictive information in Tonality for monetary policy is stronger near the lower end of the distribution of funds rate forecast errors, consistent with results for GDP and unemployment forecasts. The results find evidence of the asymmetry for 2-quarter forecast errors, but not in the case of 4-quarter forecast errors.

C. Tonality as a Predictor of Stock Returns

[17] The logic for such a connection between the Greenbook forecasts of the federal funds rate and Greenbook forecasts for unemployment seems identical; 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 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] The optimal revision inferred from the regression estimate is $(1+\beta)$*revision. A coefficient of about 1 on forecast revision suggests that revisions tend to be half their optimal size.

The evidence from the previous two sections indicating that economists' forecasts, both inside and outside of the Federal Reserve, do not contain all the information embedded in Greenbook Tonality. In addition, Tonality helps to predict errors in the publicly available forecasts of the monetary policy rate; moreover, the direction of the policy rate prediction is consistent with the direction of economic forecast errors. These results beg the question: does Tonality contain information that is not reflected in asset market prices as well? If so, then one might expect, for instance, that Greenbook Tonality could also help to predict stock market performance. In what follows, we test whether Tonality has predictive power for stock returns over the roughly 3, 6, and 12-month periods that begin the day after FOMC monetary policy announcements. Here we consider only a brief foray into tests of stock return predictability as a straightforward extension of that well-trod literature.[19]

The precise dating of the periods over which we test for return predictability is determined by FOMC dates; in each case, the period starts the day after the current-period policy announcement, and it ends on the day of a future post-meeting policy announcement. For most of the sample, the endpoints of the prediction periods correspond to the FOMC announcement days that follow the 2nd prospective meeting (about three months hence), the 4th prospective meeting (six months hence) and the 8th prospective meeting (a year hence). 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. Prediction regressions are estimated over the full sample.

Table 12 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 both the in-sample adjusted R-squared and an out-of-sample R-squared, 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 at the 6-month horizon is about double that for the 3-month horizon, and is somewhat larger again for 12-month returns. The size of these

[19] Indeed, given that we already have shown Tonality helps predict some innovations to Fed funds rates, the implications for bond return forecasting seem potentially quite interesting and deserving of careful attention, which we reserve for future study.

effects are fairly substantial. An increase in Trend Tonality of one—which amounts to roughly 1.5 standard deviations—presages a 3.6 percent higher return over the subsequent 6 months (or 4-meeting period). Although not shown, regressions that also include Tonality Shock, find it has no predictive content, consistent with its irrelevance for economic outcomes.

The adjusted R-squared statistics for the 3-month, 6-month, and 12-month horizons, are 2.1, 4.1 and 5.4, respectively, which are fairly sizable compared with most stock return predictive regressions in the literature for example (Welch and Goyal 2008). The out-of-sample R^2 statistics are also positive, in notable contrast with many out-of-sample predictive regressions. If an investor were able to take advantage of such information in real time, the gain would be economically meaningful. Using the evaluation framework of Campbell and Thompson (2007), for instance, suggests this would boost expected 6-month returns by 6.1 percent.[20]

The most natural interpretation for Tonality's predictive value is that Tonality contains information not fully reflected in stock prices at the time Greenbook is produced, but which is revealed to investors over subsequent quarters. The news of a stronger economy that higher Tonality predicates would presumably be accompanied by news of stronger corporate cash flows and perhaps a decline in risk premiums. On the contrary, it seems unintuitive and implausible to interpret Tonality's effect arising from its being a proxy for the market's risk premium; that would have the odd implication that investors demand a lower risk premium when Greenbook sentiment is more negative. Moreover, the argument that Tonality embeds information that is not reflected in stock prices is consistent with the fact that this sentiment is not publicly observable. (Indeed, it is arguable that, at the time, even Fed staff probably was not fully cognizant of the sentiment embedded in Trend Tonality.)

[20] Following Campbell and Thompson (2007), framework for gauging economic significance for a risk-averse investor, the risky asset return is 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 6.1 percent.

While we argue that Tonality is unlikely to be a proxy for the risk premium, it could well be correlated with the risk premium. If so, the prediction regression would be better specified if we could control for the market risk premium at the time of Greenbook production. One natural proxy for investors' risk premium is the expectation for current-quarter unemployment, which was shown to be correlated with Trend Tonality (table 1). Unlike Tonality, however, the Fed forecast for Current Unemployment is practically observable to the investing public, and it has a correlation of 99% with the analogous and publicly observable Blue Chip forecast. And current unemployment should be a good measure of business-cycle-driven variation in the equity risk premium to the extent that risk aversion or perceived risk are linked to employment prospects. Indeed, the perceived-risk interpretation is invoked by Schmidt (2016) as the rationale behind the return predictability he documents for initial unemployment claims, an economic statistic that is highly correlated with the Greenbook forecast of Current Unemployment.

As shown in 4th-6th columns, when Current Unemployment is added to our regression, the marginal predictive power of Trend Tonality rises, with larger positive coefficients and stronger statistical significance. Moreover, Current Unemployment appears to be an important predictor in its own right, with a significant positive coefficient, consistent with the interpretation that it serves as a proxy for the time-varying risk premium. The R-squared in each of the three regressions also rises substantially, while the effects on out-of-sample R-squared statistics are mixed. All told, our interpretation of Tonality as private information is bolstered by our finding that, when we control for time-varying risk aversion, Tonality's predictive ability seems to improve. Taken together, our evidence stacks in favor of Tonality as private information about non-modal outcomes.

Given the extensive literature on predictors of stock returns, and the lack of attention the unemployment rate has received in the return prediction literature (until Schmidt, 2016), it seems surprising that the current-quarter unemployment rate would show up here as a strong predictor of excess returns. However, its strength as a predictor here appears to owe to its complementarity with Tonality. If we also include in these regression other standard predictors, in particular, the dividend yield, as shown in the last three columns, Current Unemployment is no longer significant, while the predictive value of Trend Tonality remains robust.

A final question we consider is whether the predictive effects of Tonality for stock returns are strongest toward the lower side of potential outcomes—low to negative returns—consistent with its signal for GDP and unemployment. Table 13 provides quantile regression estimates for 3-month, 6-month and 12-month excess returns, conditioned only on Trend Tonality. At all three horizons, we find Trend Tonality to have its largest predictive effects for returns toward the lower tail of the return distributions, mirroring our results for macroeconomic predictability.

D. Is Greenbook Tonality Communicated to the Public?

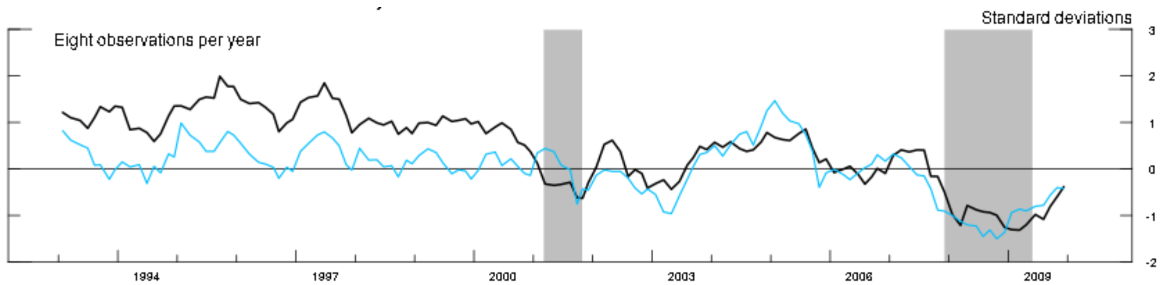
To gauge the extent to which the sentiment of the Greenbook narrative is transmitted to the public through FOMC communications, we measure the Tonality of the two regular communications issued to the public, (i) FOMC statements and (ii) minutes of the FOMC meetings. In February 1993 the committee began 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.

For each set of communications, Tonality is measured by counting positive and negative word usage in those documents and normalizing using the analogous tdf-if routine used in our analysis of the Greenbooks. The resultant time series for statement Tonality is uncorrelated with Greenbook Tonality (0.04 for full sample, same as the post-May 1999 sample). In contrast, the correlation of 0.50 between Minutes Tonality and Greenbook Tonality would appear to be quite substantial. Constructing Trend Minutes Tonality, we find its correlation with the analogous Trend Tonality for Greenbook to be even higher, at 0.74. As shown in figure 6, those two measure of sentiment look quite similar, and would appear even more so if not for their divergent trends in early 2001.

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 statistical 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 four-quarter-ahead funds rate policy and for stock returns carries through to Minutes Tonality.

Figure 6: Minutes versus Greenbook Trend Tonality



Note: Shaded regions represent NBER-dated recessions. The black line is the Greenbook Trend 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.

VII. Summary, Interpretation, and Conclusions

The predictive contribution 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 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.

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, while notable in its own right, is arguably not entirely surprising once we have established its ability to predict unexpected economic growth. Given that greater downside risks signaled by Tonality predicts lower-than-average returns, the time varying expected return documented here would not seem to reflect compensation for risk. Rather, these results suggest that equity prices do not contemporaneously 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 information that forecasters are relaying along with the quantitative point forecasts when examining forecast effectiveness or how economic agents update their beliefs. Doing so will require preserving (and in some cases) obtaining the narrative accompanying the forecasts. Quantile regressions for forecast errors seem to indicate that the information in that narrative may be disproportionately focused on the likelihood of negative tail outcomes. While the paper shows that the narrative that accompanies the Fed's economic forecast is informative in and of itself, it leaves an important question unanswered – is the narrative of other economic forecasters similarly informative or is the Federal Reserve's staff forecast special in this regard? Finally, the paper uses a relatively coarse measure of textual information. Deeper and more targeted textual analysis could lead to more insight into the economic forecasting process.

VIII. References

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

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. 2018. *How news and its context drive risk and returns around the world.* Working Paper 24430, Boston: National Bureau of Economic Research, 1-77.

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.

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.

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.

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

Gregory, Karl B, Soumendra N Lahiri, and and Dan J Nordman. forthcoming. "A Smooth Block Bootstrap for Quantile Regression with Time Series." *Annals of Statistics* 1-30.

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.

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.

Lettau, Martin, and Sydney Ludvigson. 2001. "Consumption, Aggregate Wealth, and Expected Stock Returns." *The Journal of Finance* 56 (3): 815-849.

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. 2012. "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.

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.

Shapiro, Adam H., Moritz Sudhof, and Daniel J. Wilson. 2017. "Measuring News Sentiment." *Federal Reserve Bank of San Francisco Working Paper* 1-32.

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.

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.

Thorsrud, Leif Anders. 2016. *Nowcasting using news topics. Big Data versus big bank.* Working Paper 20, Oslo: Norges Bank Research.

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.

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.

Table 1: Pearson Correlation of Text Tonality with Greenbook point forecast variables

	Tonality	Trend Tonality
Current GDP Growth Forecast	0.17***	0.29***
GDP 2-Qtr Growth Forecast	0.22***	0.29***
GDP 2-Qtr Forecast Revision	0.26***	0.20***
Current Unemployment Forecast	-0.07	-0.24***
Unemployment 2-Qtr Forecast	-0.33***	-0.46***
Unemployment 2-Qtr Forecast Revision	-0.27***	-0.25***
Current Inflation Forecast	-0.32***	-0.43***
Inflation 2-Qtr Forecast	-0.33***	-0.49***
Inflation 2-Qtr Forecast Revision	-0.10*	-0.13**
Uncertainty	-0.16***	-0.14***

Notes: Current GDP growth is the RGDP growth rate for the current quarter as expected by the staff forecast in the Greenbook, GDP 2-Qtr is the cumulative 2-quarter GDP growth between the previous quarter and 2-quarters later. GDP revision is the revision to the GDP 2-Qtr growth forecast from previous Greenbook to the current Greenbook. Unemployment and Inflation forecast and revision variables are similarly defined with respect to the change in unemployment rate and cumulative change in inflation. *p<0.1; **p<0.05; ***p<0.01

Table 2: Pearson Correlations among Greenbook forecast variables

	GDP Forecast	Unemp Forecast	Infl Forecast	GDP Rev	Unemp Rev	Infl Rev
GDP Forecast						
Unemp Forecast	-0.86***					
Infl Forecast	-0.29***	0.26***				
GDP Rev	0.33***	-0.32***	0.00			
Unemp Rev	-0.36***	0.47***	-0.02	-0.69***		
Infl Rev	-0.01	0.03	0.32***	0.04	-0.09	
Uncertainty	-0.14***	0.20***	0.01	-0.20***	0.17***	0.12**

Notes: To ease reading, we provide only the lower triangular matrix. *p<0.1; **p<0.05; ***p<0.01

Table 3: Greenbook Text Tonality and the Dual Mandate

	Tonality			Trend Tonality	
	Full Sample	Up to 1991-10-30	Post 1991-10-30	Up to 1991-10-30	Post 1991-10-30
Inflation Forecast	−0.217***	−0.165***	0.496***	−0.139***	0.427***
	(0.042)	(0.054)	(0.172)	(0.022)	(0.100)
Unemployment Forecast	−0.347***	−0.122	−1.228***	−0.185***	−1.116***
	(0.077)	(0.081)	(0.152)	(0.033)	(0.088)
Intercept	0.546***	0.198	0.066	0.135**	0.135
	(0.102)	(0.161)	(0.220)	(0.066)	(0.128)
Observations	358	213	145	213	145
Adjusted R^2	0.148	0.071	0.367	0.356	0.586
Residual Std. Error	0.953	0.847	0.855	0.346	0.496
F Statistic	31.971***	9.150***	42.655***	59.687***	103.065***

Notes: *p<0.1; **p<0.05; ***p<0.01

Table 4: Regressions Predicting Cumulative GDP Growth

	Quarters Ahead					
	2	4	2	4	2	4
Staff Forecast	0.96***	0.80***	0.93***	0.74***	0.90***	0.69***
	(0.11)	(0.11)	(0.10)	(0.11)	(0.10)	(0.11)
Tonality			0.26**	0.61***		
			(0.11)	(0.19)		
Trend Tonality					0.47**	1.10***
					(0.23)	(0.39)
Tonality Shock					0.04	0.08
					(0.11)	(0.18)
Intercept	0.04	0.49	0.10	0.68	0.15	0.79
	(0.33)	(0.57)	(0.31)	(0.53)	(0.30)	(0.52)
P(Forecast = 1)	0.69	0.07	0.47	0.01	0.32	0.01
Observations	358	318	358	318	358	318
Residual Std. Error	1.42	2.10	1.40	2.02	1.39	1.97
Adjusted R^2	0.55	0.40	0.57	0.45	0.58	0.48
Out-of-sample R^2	0.41	0.34	0.42	0.39	0.41	0.41
Note:	*p<0.1; **p<0.05; ***p<0.01					

Notes: Estimates from the regression of 2- and 4- quarter cumulative RGDP growth on Fed Staff forecast of cumulative RGDP growth, and Tonality (or Trend and Shock components of Tonality). Table shows estimates between January 1972 to December 2009. Standard errors shown below coefficient estimates are corrected for autocorrelation for (2*k +1) lags for k quarter out forecast error regression using the automatic bandwidth selection procedure described in (Newey and West 1994).The out-of-sample R^2 are calculated over the period that begins 64 meetings into the start of the sample through December 2009. $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table 5: Regressions Predicting Unemployment Change

	Quarters Ahead					
	2	4	2	4	2	4
Staff Forecast	1.12***	1.19***	1.09***	1.10***	1.04***	1.02***
	(0.11)	(0.15)	(0.11)	(0.14)	(0.10)	(0.13)
Tonality			−0.06*	−0.21***		
			(0.04)	(0.08)		
Trend Tonality					−0.16**	−0.43***
					(0.07)	(0.15)
Tonality Shock					0.03	−0.02
					(0.03)	(0.08)
Intercept	−0.08	−0.07	−0.07	−0.04	−0.06	−0.01
	(0.05)	(0.11)	(0.05)	(0.11)	(0.05)	(0.12)
Observations	358	318	358	318	358	318
P(Forecast = 1)	0.29	0.22	0.39	0.48	0.66	0.90
Residual Std. Error	0.57	0.88	0.56	0.85	0.56	0.83
Adjusted R^2	0.64	0.55	0.65	0.58	0.65	0.59
Out-of-sample R^2	0.64	0.51	0.64	0.54	0.64	0.53
Note:	*p<0.1; **p<0.05; ***p<0.01					

Notes: Estimates from the regression of 2- and 4- quarter unemployment change on Fed Staff forecast of unemployment change, and Tonality (or Trend and Shock components of Tonality). Table shows estimates between January 1972 to December 2009. Standard errors shown below coefficient estimates are corrected for autocorrelation for (2*k +1) lags for k quarter out forecast error regression using the automatic bandwidth selection procedure described in (Newey and West 1994).The out-of-sample R^2 are calculated over the period that begins 64 meetings into the start of the sample through December 2009. *p<0.1; **p<0.05; ***p<0.01

Table 6: Regressions Predicting Cumulative Inflation Growth

Quarters Ahead	2	4	2	4	2	4
Panel A. 1972-1991						
Staff Forecast	0.80***	0.87***	0.77***	0.81***	0.66***	0.72***
	(0.10)	(0.15)	(0.10)	(0.16)	(0.12)	(0.24)
Tonality			−0.14	−0.39		
			(0.11)	(0.30)		
Trend Tonality					−0.72*	−1.11
					(0.44)	(1.25)
Tonality Shock					0.08	−0.16
					(0.10)	(0.17)
Intercept	0.61**	0.80	0.65**	1.06	0.79**	1.44
	(0.30)	(0.97)	(0.30)	(1.06)	(0.33)	(1.39)
Observations	213	173	213	173	213	173
P(Forecast = 1)	0.05	0.39	0.02	0.24	0.01	0.24
Residual Std. Error	1.01	1.92	1.00	1.90	0.98	1.88
Adjusted R^2	0.46	0.47	0.46	0.48	0.49	0.48
Out-of-sample R^2	0.42	0.60	0.43	0.59	0.46	0.57
Panel B. 1991-2009						
Staff Forecast	0.31	0.07	0.25	−0.16	0.16	−0.30
	(0.21)	(0.20)	(0.21)	(0.20)	(0.23)	(0.27)
Tonality			0.08	0.40**		
			(0.12)	(0.18)		
Trend Tonality					0.25	0.71**
					(0.21)	(0.35)
Tonality Shock					−0.12	0.06
					(0.13)	(0.12)
Intercept	0.88***	2.32***	0.92***	2.67***	0.95***	2.86***
	(0.30)	(0.53)	(0.27)	(0.43)	(0.26)	(0.48)
Observations	145	145	145	145	145	145
P(Forecast = 1)	0.00	0.00	0.00	0.00	0.00	0.00
Residual Std. Error	0.77	1.09	0.77	1.01	0.75	0.98
Adjusted R^2	0.02	−0.005	0.03	0.12	0.06	0.19
Out-of-sample R^2	0.69	0.80	0.69	0.79	0.66	0.78

Notes: Estimates from the regression of 2- and 4- quarter cumulative inflation growth on Fed Staff forecast of cumulative inflation growth, and Tonality (or Trend and Shock components of Tonality). Panel A shows estimates between 1970 to October 1991; Panel B shows estimates after December 1992 to December 2009. Standard errors shown below coefficient estimates are corrected for autocorrelation for (2*k +1) lags for k quarter out forecast error regression using the automatic bandwidth selection procedure described in (Newey and West 1994).The out-of-sample R^2 are calculated over the period that begins 64 meetings into the start of the sample through December 2009. *p<0.1; **p<0.05; ***p<0.01

Table 7: Forecast Regressions with Additional Variables

	GDP				Unemployment			
Quarters Ahead	2	4	2	4	2	4	2	4
Forecast	0.89***	0.72***	0.86***	0.67***	1.09***	1.08***	1.03***	0.98***
	(0.10)	(0.11)	(0.10)	(0.12)	(0.11)	(0.15)	(0.10)	(0.14)
Tonality	0.15	0.49***			−0.03	−0.16**		
	(0.11)	(0.17)			(0.03)	(0.07)		
Trend Tonality			0.39*	0.96***			−0.14**	−0.39***
			(0.21)	(0.35)			(0.06)	(0.14)
Tonality Shock			−0.10	−0.03			0.07*	0.05
			(0.12)	(0.19)			(0.04)	(0.08)
Uncertainty	−0.07	−0.20	−0.07	−0.21	0.01	0.07	0.02	0.08
	(0.11)	(0.19)	(0.11)	(0.21)	(0.04)	(0.06)	(0.04)	(0.06)
Staff Revision	0.32	0.08	0.35	0.13	0.004	−0.02	0.04	0.04
	(0.25)	(0.29)	(0.25)	(0.28)	(0.16)	(0.26)	(0.16)	(0.26)
Recent Stock Return	0.06***	0.05***	0.06***	0.05***	−0.03***	−0.03***	−0.03***	−0.03***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
Intercept	0.16	0.74	0.22	0.87	−0.06	−0.02	−0.04	0.01
	(0.31)	(0.55)	(0.29)	(0.54)	(0.05)	(0.11)	(0.05)	(0.12)
Observations	355	315	355	315	355	315	355	315
P(Forecast = 1)	0.29	0.01	0.16	0.00	0.40	0.59	0.77	0.89
Residual Std. Error	1.36	1.94	1.34	1.89	0.55	0.83	0.54	0.81
Adjusted R^2	0.59	0.47	0.60	0.50	0.66	0.59	0.67	0.61
Out-of-sample R^2	0.37	0.38	0.35	0.40	0.62	0.54	0.62	0.55

Notes: Estimates from the regression of 2- and 4- quarter cumulative RGDP growth and unemployment change on their respective Fed Staff forecasts, Tonality (or Trend and Shock components of Tonality), uncertainty count and respective forecast revision. Table shows estimates between January 1972 to December 2009. Standard errors shown below coefficient estimates are corrected for autocorrelation for (2*k +1) lags for k quarter out forecast error regression using the automatic bandwidth selection procedure described in (Newey and West 1994).The out-of-sample R^2 are calculated over the period that begins 64 meetings into the start of the sample through December 2009. $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table 8a: Quantile Regressions Predicting GDP Forecast Errors

	2-Qtr Errors				4-Qtr Errors			
	Tonality	Psuedo R^2	Trend Tonality	Psuedo R^2	Tonality	Psuedo R^2	Trend Tonality	Psuedo R^2
Q_{90}	-0.23*	0.01	-0.43***	0.05	0.21	0.00	-0.20	0.01
	(0.13)		(0.14)		(0.17)		(0.20)	
Q_{75}	0.07	0.00	0.10	0.00	0.27	0.02	0.40*	0.02
	(0.11)		(0.15)		(0.17)		(0.23)	
Q_{50}	0.21**	0.02	0.43**	0.03	0.38**	0.02	0.90***	0.05
	(0.10)		(0.17)		(0.18)		(0.26)	
Q_{25}	0.31***	0.03	0.66***	0.06	0.52***	0.05	1.15***	0.09
	(0.10)		(0.16)		(0.16)		(0.25)	
Q_{10}	0.44***	0.06	0.84***	0.13	0.75***	0.08	1.62***	0.13
	(0.10)		(0.15)		(0.15)		(0.26)	
$P(Q_{50} = Q_{90})$	0.043		0.007		0.273		0.04	
$P(Q_{50} = Q_{10})$	0.127		0.049		0.211		0.191	

Notes: Estimates from the quantile regressions of 2- and 4- quarter cumulative RGDP growth forecast errors on Tonality (or Trend and Shock components of Tonality). Table shows the estimates for July 1972 to December 2009 in columns 1, 3, 5 and 7 and Pseudo R^2 statistic as described in (Koenker and Machado 1999) in columns 2, 4, 6 and 8. Trend and Shock components of Tonality are derived by constructing an exponentially weighted moving average of Tonality. Shock components are present in the estimations shown in columns 2, 4 and 6, but are omitted from the table. Quantile regression tests are performed using a smooth block bootstrap as described in (Gregory, Lahiri and Nordman 2018).*p<0.1; **p<0.05; ***p<0.01

Table 8b: Quantile Regressions Predicting Unemployment Forecast Errors

	2-Qtr Errors				4-Qtr Errors			
	Tonality	Psuedo R^2	Trend Tonality	Psuedo R^2	Tonality	Psuedo R^2	Trend Tonality	Psuedo R^2
Q_{90}	-0.22***	0.07	-0.42***	0.19	-0.47***	0.13	-0.82***	0.22
	(0.05)		(0.05)		(0.07)		(0.10)	
Q_{75}	-0.14***	0.04	-0.27***	0.10	-0.28***	0.06	-0.49***	0.10
	(0.04)		(0.06)		(0.06)		(0.09)	
Q_{50}	-0.03	0.00	-0.04	0.00	-0.13**	0.03	-0.25***	0.04
	(0.04)		(0.06)		(0.05)		(0.07)	
Q_{25}	0.02	0.00	0.06	0.00	-0.04	0.00	-0.13*	0.01
	(0.04)		(0.05)		(0.05)		(0.07)	
Q_{10}	0.15***	0.03	0.23***	0.04	0.02	0.00	0.05	0.00
	(0.05)		(0.06)		(0.05)		(0.08)	
$P(Q_{50} = Q_{90})$	0.003		0.000		0.004		0.002	
$P(Q_{50} = Q_{10})$	0.082		0.050		0.231		0.076	

Notes: Estimates from the quantile regressions of 2- and 4- quarter unemployment change forecast errors on Tonality (or Trend and Shock components of Tonality). Table shows the estimates for July 1972 to December 2009 in columns 1, 3, 5 and 7 and Pseudo R^2 statistic as described in (Koenker and Machado 1999) in columns 2, 4, 6 and 8. Trend and Shock components of Tonality are derived by constructing an exponentially weighted moving average of Tonality. Shock components are present in the estimations shown in columns 2, 4 and 6, but are omitted from the table. Quantile regression tests are performed using a smooth block bootstrap as described in (Gregory, Lahiri and Nordman 2018).*p<0.1; **p<0.05; ***p<0.01

Table 9: Regressions Predicting Bluechip 4-Qtr Forecast Errors

	GDP				Unemployment			
Quarters Ahead	2	4	2	4	2	4	2	4
Trend Tonality	0.39	1.15^{***}	0.35	1.11^{***}	-0.28^{***}	-0.67^{***}	-0.20^{**}	-0.54^{***}
	(0.24)	(0.38)	(0.23)	(0.38)	(0.11)	(0.19)	(0.09)	(0.17)
Tonality Shock	0.05	0.01	0.02	−0.03	0.03	−0.04	0.06	0.003
	(0.11)	(0.15)	(0.11)	(0.14)	(0.05)	(0.08)	(0.04)	(0.07)
Revision			0.46	−0.20			0.74^{***}	0.84^{***}
			(1.42)	(2.70)			(0.28)	(0.29)
BC Stock Return			0.02	0.03			-0.01^{*}	-0.03^{***}
			(0.01)	(0.02)			(0.01)	(0.01)
Intercept	0.01	−0.40	0.01	−0.42	0.03	0.22	0.01	0.20
	(0.18)	(0.29)	(0.17)	(0.29)	(0.08)	(0.17)	(0.07)	(0.15)
Observations	239	237	239	237	239	237	239	237
Residual Std. Error	1.16	1.70	1.16	1.70	0.51	0.82	0.48	0.79
Adjusted R^2	0.05	0.19	0.05	0.19	0.13	0.25	0.23	0.31
Out-of-sample R^2	0.092	0.135	0.090	0.106	0.143	0.100	0.251	0.164

Notes: Estimates from the regression of 4- quarter cumulative GDP growth and unemployment rate change errors on Trend and Shock components of Tonality for January 1980 to 2009. The first two columns show GDP growth rate regression estimates, the next two show change in unemployment rate regression estimates. Standard errors shown below coefficient estimates are corrected for autocorrelation for (2*k +1) lags for k quarter out forecast error regression using the automatic bandwidth selection procedure described in (Newey and West 1994). The out-of-sample R^2 are calculated over the period that begins 64 meetings into the start of the sample through December 2009. *p<0.1; **p<0.05; ***p<0.01

Table 10: Quantile Regressions Predicting 4-Qtr Bluechip Errors: 1978*-2009

	RGDP				Unemployment			
	Tonality	Psuedo R^2	Trend Tonality	Psuedo R^2	Tonality	Psuedo R^2	Trend Tonality	Psuedo R^2
Q_{90}	-0.20	0.01	0.41**	0.05	-0.54***	0.32	-1.04***	0.36
	(0.14)		(0.18)		(0.06)		(0.12)	
Q_{75}	0.05	0.01	0.65***	0.06	-0.42***	0.18	-0.69***	0.22
	(0.16)		(0.23)		(0.07)		(0.13)	
Q_{50}	0.46**	0.05	1.12***	0.08	-0.15**	0.06	-0.40***	0.12
	(0.20)		(0.29)		(0.07)		(0.10)	
Q_{25}	0.88***	0.10	1.29***	0.13	-0.04	0.02	-0.21***	0.08
	(0.17)		(0.28)		(0.06)		(0.08)	
Q_{10}	0.70***	0.15	1.50***	0.23	0.09*	0.03	0.07	0.01
	(0.15)		(0.26)		(0.06)		(0.09)	
$P(Q_{50} = Q_{90})$	0.100		0.284		0.000		0.004	
$P(Q_{50} = Q_{10})$	0.454		0.302		0.020		0.048	

Notes: Estimates from the quantile regressions of 4- quarter RGDP and unemployment change forecast errors on Tonality (or Trend and Shock components of Tonality). Table shows the estimates for January 1980 to December 2009 in columns 1, 3, 5, and 7 and Pseudo R^2 statistic as described in (Koenker and Machado 1999) in columns 2, 4, 6, and 8. Trend and Shock components of Tonality are derived by constructing an exponentially weighted moving average of Tonality. Shock components are present in the estimations, but are omitted from the table. Quantile regression tests are performed using a smooth block bootstrap as described in (Gregory, Lahiri and Nordman 2018).*p<0.1; **p<0.05; ***p<0.01

Table 11: Regressions Predicting Bluechip Fed Funds Errors: 1986-2009

Quarters Ahead	2	4	2	4
Trend Tonality	0.25^{**}	0.59^{***}	0.15^{*}	0.40^{**}
	(0.10)	(0.22)	(0.09)	(0.20)
Tonality Shock	0.11	0.10	0.06	0.04
	(0.09)	(0.18)	(0.08)	(0.16)
Revision			0.59^{***}	1.04^{***}
			(0.18)	(0.29)
Recent Stock Return			0.01	0.01
			(0.01)	(0.02)
Intercept	-0.32^{***}	-0.75^{***}	-0.24^{**}	-0.59^{**}
	(0.11)	(0.28)	(0.10)	(0.26)
Observations	192	192	192	192
Residual Std. Error	0.70	1.30	0.68	1.26
Adjusted R^2	0.06	0.09	0.13	0.15
Out-of-sample R^2	0.097	0.117	0.174	0.150
Note:	*p<0.1; **p<0.05; ***p<0.01			

Notes: Estimates from the regression of 2- and 4- quarter Blue Chip Fed Funds Rate forecast errors on Trend and Shock components of Tonality, forecast revision and stock market return from the prior Blue Chip release. Table shows estimates between January 1986 to December 2009. Standard errors shown below coefficient estimates are corrected for autocorrelation for (2*k +1) lags for k quarter out forecast error regression using the automatic bandwidth selection procedure described in (Newey and West 1994).The out-of-sample R^2 are calculated over the period that begins 64 meetings into the start of the sample through December 2009. $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table 12: Regressions Predicting Excess S&P 500 Returns over 3-,6-,12- Months: 1970-2009

Months Ahead	3	6	12	3	6	12	3	6	12
Trend Tonality	1.74**	3.56**	5.75*	2.28***	4.45***	7.19**	2.64***	5.64***	9.51***
	(0.75)	(1.60)	(2.95)	(0.75)	(1.61)	(3.03)	(0.84)	(1.84)	(3.49)
Current Unemp				1.10***	1.81***	2.94*	0.84*	0.96	1.28
				(0.35)	(0.69)	(1.58)	(0.50)	(0.96)	(1.86)
D/P							1.57	5.22	10.15
							(1.85)	(3.61)	(6.74)
Intercept	0.03	0.10	−0.09	−6.83***	−11.19**	−18.40*	0.29	12.51	27.71
	(0.53)	(1.06)	(2.17)	(2.32)	(4.57)	(10.62)	(8.96)	(17.26)	(31.35)
Observations	358	358	358	358	358	358	358	358	358
Residual Std. Error	7.80	11.53	16.44	7.66	11.27	15.94	7.66	11.17	15.66
Adjusted R^2	0.02	0.04	0.05	0.06	0.09	0.11	0.06	0.10	0.14
Out-of-sample R^2	0.018	0.020	0.048	-0.021	0.028	0.047	-0.035	-0.007	-0.002

Notes: Returns are measured over roughly a 3-month horizon, a 6-month horizon, and a 12-month horizon, each beginning with closing prices on the current-Greenbook FOMC announcement day. For observations after 1980, the endpoints of the two prediction periods correspond to the FOMC announcement days that follow the second prospective meeting (about three months hence), the fourth prospective meeting (six months hence), and the eight prespective meeting (twelve months hence). Trend Tonality is the trend component of Tonality. The unemployment rate forecast corresponds to the quarter of the Greenbook. Standard errors shown below coefficient estimates are corrected for autocorrelation for 1, 3 and 9 lags respectively using the automatic bandwidth selection procedure described in (Newey and West 1994). The out-of-sample R^2 shows fit of S&P 500 returns from the prediction regression versus the historical mean. The out-of-sample R^2 are calculated over the period June 1975 through December 2009. *p<0.1; **p<0.05; ***p<0.01

Table 13: Quantile Regressions Predicting Excess S&P 500 3-, 6- and 12- Month Returns: 1970-2009

	3-Month		6-Month		12-Month	
	Trend Tonality	Psuedo R^2	Trend Tonality	Psuedo R^2	Trend Tonality	Psuedo R^2
Q_{90}	-0.81	0.01	-1.46	0.01	0.12	0.00
	(0.71)		(1.04)		(1.14)	
Q_{75}	0.11	0.00	0.50	0.00	2.61*	0.01
	(0.73)		(1.16)		(1.45)	
Q_{50}	1.47*	0.01	3.38***	0.02	4.06**	0.02
	(0.80)		(1.20)		(1.83)	
Q_{25}	3.44***	0.06	6.70***	0.09	9.80***	0.06
	(0.72)		(1.15)		(2.10)	
Q_{10}	4.38***	0.07	8.47***	0.11	13.86***	0.11
	(0.71)		(1.15)		(2.03)	
$P(Q_{50} = Q_{90})$	0.043		0.033		0.320	
$P(Q_{50} = Q_{10})$	0.110		0.060		0.133	

Notes: Estimates from the quantile regressions of Excess 6-Month S&P Returns on Trend Tonality Table shows the estimates for July 1970 to December 2009 in columns 1, 3, and 5 and Pseudo R^2 statistic as described in (Koenker and Machado 1999) in columns 2, 4, and 6. Trend and Shock components of Tonality are derived by constructing an exponentially weighted moving average of Tonality. Shock components are present in the estimations shown in columns 2, 4 and 6, but are omitted from the table. Quantile regression tests are performed using a smooth block bootstrap as described in (Gregory, Lahiri and Nordman 2018).*p<0.1; **p<0.05; ***p<0.01

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	

To provide some sense of the words that go into the constitution of Tonality, we provide word clouds showing the 50 most prominent positive and negative words in Greenbook during a couple different time periods. **Figure A1** shows two side-by-side word clouds for the 50 most prominent positive words in Greenbooks during two periods, 1994-1998 and 2005-2009. 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 language drift, whereby many words fall out of favor and are replaced by new ones. The most important positive word in both periods is "upward", followed closely by "positive." On the other hand, the words "favorable" and "moderation" are more prominent during 1994-1998.

Figure A2 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 "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.

Figure A1: Word cloud for fifty most positive 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.

Figure A2: Word cloud for fifty most negative words in the Greenbook.

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.

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.

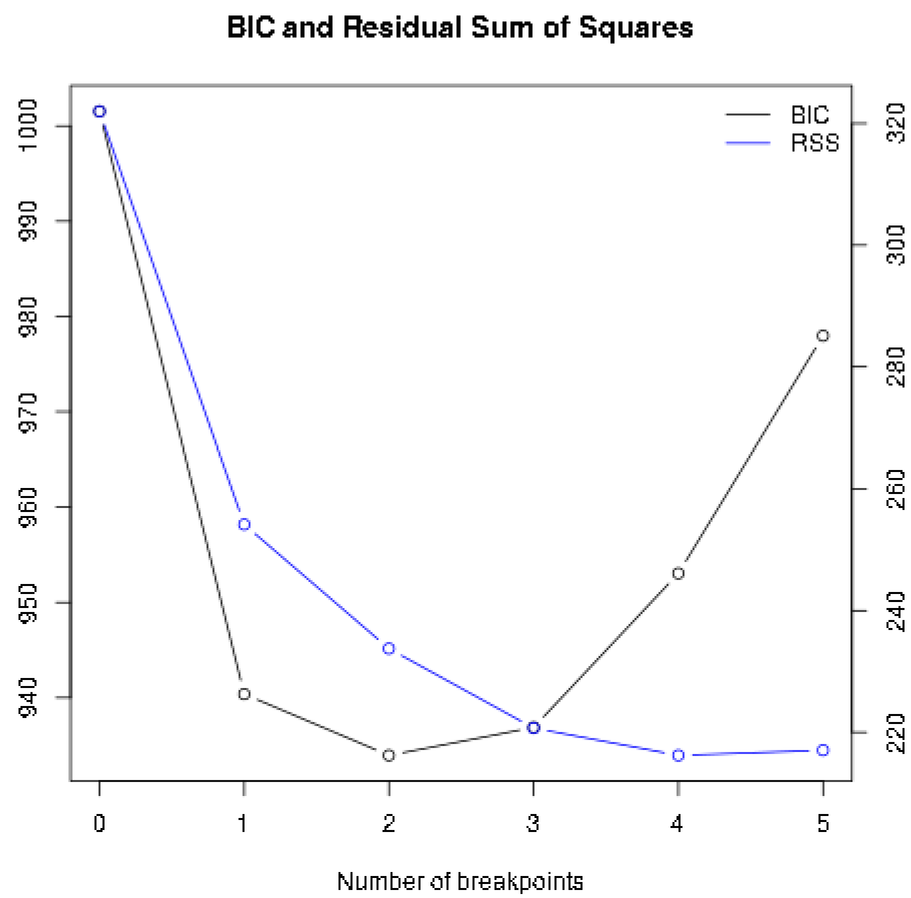
Current Unemployment$_i$ is the Staff's projection for the current unemployment rate.

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).

Appendix C: Structural break in the relationship between Tonality and Greenbook forecasts

We used the Bai and Perron (2003) test for multiple structural breaks in the econometric relationship between Tonality and Greenbook forecast variables shown in Table 2. We find strong evidence for a single break, estimated to have occurred in October, 1991. In particular, the plot below shows the Bayesian information criteria (BIC) and the residual sum of squares (RSS) as the number of breakpoints is varied between zero and five.

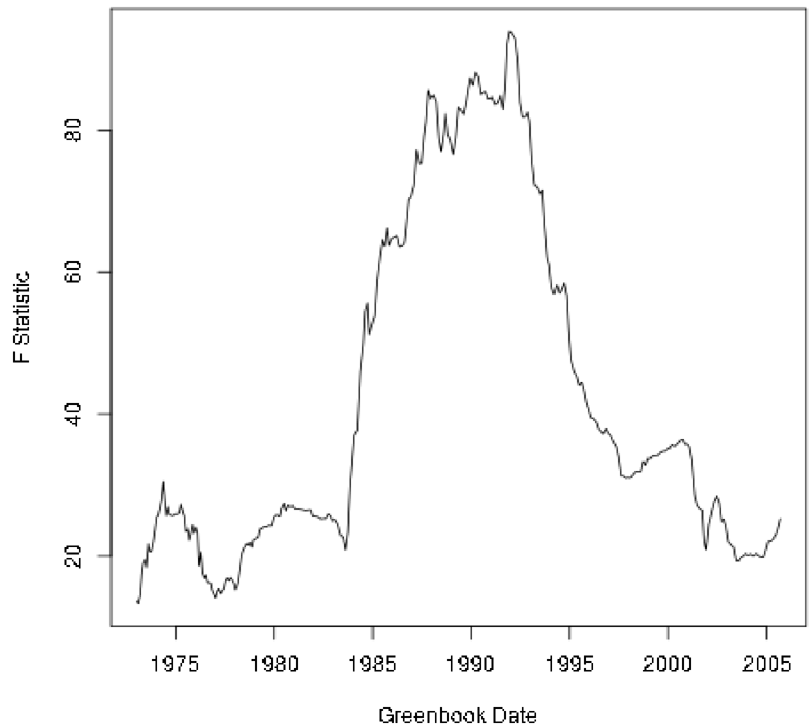
Figure C1: Number of breakpoints and model improvement



Note: The plot shows the decrease in Bayesian information criteria (BIC) and residual sum of square as we increase the number of breakpoints in the relationship between Tonality and current Greenbook point forecasts. The breakpoint corresponds to October 1991.

The test uncovers strong statistical evidence for a single break in this relationship, estimated to have occurred in October 1991. A plot of F-test values for all possible (single) breaks (Figure C2) suggests the indicated timing of the structural break is fairly definitive.

Figure C2: Chow tests for structural break in Tonality



Note: F statistics from Chow tests for structural break in regression of Tonality on numerical forecast variables