

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

**What's the Story? A New Perspective on the Value of Economic
Forecasts**

Steve Sharpe, Nitish Sinha, and Christopher A. Hollrah

2017-107

Please cite this paper as:

Sharpe, Steve, Nitish R. Sinha, and Christopher A. Hollrah (2017). "What's the Story? A New Perspective on the Value of Economic Forecasts," Finance and Economics Discussion Series 2017-107. Washington: Board of Governors of the Federal Reserve System, <https://doi.org/10.17016/FEDS.2017.107r1>.

NOTE: Staff working papers in the Finance and Economics Discussion Series (FEDS) are preliminary materials circulated to stimulate discussion and critical comment. The analysis and conclusions set forth are those of the authors and do not indicate concurrence by other members of the research staff or the Board of Governors. References in publications to the Finance and Economics Discussion Series (other than acknowledgement) should be cleared with the author(s) to protect the tentative character of these papers.

What's the Story? A New Perspective on the Value of Economic Forecasts

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

First draft: August 30, 2017
Current draft: August 01, 2018

Abstract

We apply textual analysis tools to measure the degree of optimism versus pessimism of the text that describes Federal Reserve Board forecasts published in the Greenbook. The resulting measure of Greenbook text sentiment, “Tonality,” is found to be strongly correlated, in the intuitive direction, with the Greenbook point forecast for key economic variables such as unemployment and inflation. We then examine whether Tonality has incremental power for predicting unemployment, GDP growth, and inflation up to four quarters ahead. We find it to have significant and substantive predictive power for both GDP growth and unemployment, particularly since 1991: higher (more optimistic) Tonality presages higher GDP growth and lower unemployment, relative to the Greenbook point forecasts. We then test whether Tonality helps predict monetary policy and stock returns. Higher Tonality has some power to predict tighter than forecasted monetary policy, while it has substantial power for predicting higher 3-month, 6-month, and 12-month stock market returns.

JEL codes: C53, E17, E27, E37, E52, G40.

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. Also, many thanks to Jeremy Rudd for his help in developing the wordlists.

What's the Story? A New Perspective on the Value of Economic Forecasts

Abstract

We apply textual analysis tools to measure the degree of optimism versus pessimism of the text that describes Federal Reserve Board forecasts published in the Greenbook. The resulting measure of Greenbook text sentiment, “Tonality,” is found to be strongly correlated, in the intuitive direction, with the Greenbook point forecast for key economic variables such as unemployment and inflation. We then examine whether Tonality has incremental power for predicting unemployment, GDP growth, and inflation up to four quarters ahead. We find it to have significant and substantive predictive power for both GDP growth and unemployment, particularly since 1991: higher (more optimistic) Tonality presages higher GDP growth and lower unemployment, relative to the Greenbook point forecasts. We then test whether Tonality helps predict monetary policy and stock returns. Higher Tonality has some power to predict tighter than forecasted monetary policy, while it has substantial power for predicting higher 3-month, 6-month, and 12-month stock market returns.

I. Introduction

Over the years, many researchers and market participants have questioned the value of economic forecasts, characterizing what seems to be a less than stellar record. 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). In December 2010, for instance, the document containing the staff forecast was over 100 pages long, with tables detailing forecasts for about 50 U.S. macroeconomic data series, plus dozens of additional series detailing forecasts of the federal budget, credit flows across sectors, as well as GDP and inflation for major foreign countries and regions. This paper provides a new perspective on the value of forecasts, which can help explain why financial market participants and policy makers continue to pay for them.

In the academic literature, economic forecasts by the Federal Reserve staff as well as those from the private sector and academia 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 estimates of inflation and/or GDP growth, which are usually interpreted as either modal or mean predictions. Consequently, these studies ignore a major element of the forecasters’ product, the narratives in which the quantitative forecasts are embedded. Such narratives tend to give a flavor of the range of plausible outcomes or characterize the direction of likely risks to forecasts. This shortcoming of traditional research on forecast efficacy is not surprising, as quantitative forecasts have been conveniently catalogued for decades. However, it is plausible that policymakers and investors who pay for these forecasts

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.

draw significant value from the narratives that accompany individual forecasts, and new methods of text analysis offer the opportunity to explore this angle.

Our study breaks new ground by applying tools from the emerging literature on textual analysis in an attempt to gauge a key dimension of the information conveyed in the narratives that accompany forecasts. To do so, we focus on Federal Reserve Board forecasts published in the Greenbook. In particular, we quantify the degree of optimism versus pessimism embedded in the Greenbook text, 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-social dictionary, which is then fine-tuned by excluding words that have special meaning in an economic forecasting context, such as “demean” and “interest.” The resulting measure of Greenbook text sentiment is strongly correlated with the point forecasts for key economic variables in the Greenbook, specifically, forecasts for GDP growth, unemployment and inflation providing some assurance that our measure of sentiment does reflect key factors that should influence the narrative in the Greenbook.

We then examine whether the resulting measure of optimism has power, over and above numerical forecasts, for predicting key macroeconomic quantities—namely unemployment, GDP growth, and inflation. We consider horizons ranging from one quarter to four quarters ahead. In short, we find that Tonality has significant predictive power, particularly for the cumulative change in unemployment and GDP growth over the subsequent four-quarter horizon. During the post-1991 period, when the power of Tonality is most pronounced, including it in the information set improves the out-of-sample R² for the four-quarter-ahead GDP growth forecast from 24% to 37%; similarly, the out-of-sample R² for the four-quarter-ahead unemployment rate forecast is increased from 45% to 50%.

In light of the predictive power of Tonality for economic activity, we test the logical corollary, which seems particularly relevant given the identity of the forecaster: does Tonality of the text help to predict monetary policy surprises? For this test, we use two alternative measures of monetary policy expectations: (i) the Fed staff’s Greenbook funds rate forecasts and, (ii) the consensus Blue Chip forecasts for the Fed funds rate. Indeed, using either measure as the benchmark forecast, we find that Tonality has significant predictive power for monetary policy.

In particular, a more optimistic tone presages a higher than anticipated Fed funds rate in the near-term and up to four quarters ahead.

We then examine whether Tonality can predict stock returns over similar horizons, and these tests for the predictive content of Tonality turn up fairly striking results. Tonality has substantial power for predicting excess returns on stocks over the 3-, 6- and 12-month holding periods that follow Greenbook's distribution to policymakers. Moreover, that predictability holds up in out-of-sample tests much better than conventional (observable) conditioning variables. And predictability is even stronger when we control for the current-quarter forecast of unemployment, a proxy for the investors' risk premium. The positive coefficient on Tonality is consistent with the interpretation that its predictive power arises from its ability to predict news that investors will receive. In particular, subsequent news of a stronger economy raises expected profits and dividends and presumably lowers investors' risk premiums, all of which would provide a boost to stock values. Thus, Tonality would appear to impound Greenbook authors' private information or public information that is underappreciated by investors.

Finally, we consider why Tonality contains information absent from the numerical forecast. Our finding that it similarly helps to predict Blue Chip forecasts, and that it predicts stock returns, indicates that it does not owe to a purely internal Fed dynamic. One plausible explanation rests on the presumption that numerical forecasts, in this case both the Greenbook forecast and the Blue Chip median, are modal forecasts, whereas Tonality could reflect the ebb and flow of information about various non-modal or even tail outcomes. Indeed, if forecasters are cognizant of risks to the forecast, the text could reflect information about tail outcomes. We find some evidence for this conjecture. In particular, Tonality helps predict recessions in a simple probit model that controls for the Greenbook's numerical forecast. Relatedly, in a quantile regression specification, we find that Tonality is more informative about the higher quantiles of unemployment forecast errors.

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. Among the most widely cited text-as-data studies in economics is a study by Baker, Bloom and Davis (2016), who create measures of government economic and monetary policy uncertainty by measuring the usage of language in

newspaper articles on the subject. The approach used in our analysis also has close parallels to recent studies that examine how the tone of newspaper articles helps explain or predict stock market returns beginning with Tetlock (2007), using techniques that continue to be elaborated upon, for instance by Heston and Sinha (2017) and Calomiris and Mamaysky (2018). A related line of studies examine how the text in company earnings reports or equity analyst updates helps explain the company's stock price responses to earnings forecast revisions (Asquith, Mikhail and Au 2005). Shapiro, Sudhof and Wilson (2017) find 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, relatedly Thorsrud (2016) uses news topics to construct a “nowcast” of the Norwegian economy.

Perhaps even more closely 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.

In section II, we describe how we measure Tonality and explore how it co-varies with Fed staff's key quantitative forecasts. In section III, we examine whether Tonality can predict macroeconomic conditions. In section IV, we examine the relationship between Tonality and monetary policy. Section V explores relationship between Tonality and future stock prices. Section VI examines three extensions – first, Tonality as a predictor of recessions, second whether Tonality contains information about extreme states of the economy as well stock returns

in a quantile regression setting, and finally information content of positive and negative Tonality. Section VII concludes.

II. Measurement of Tonality in Greenbook Text

A. Measuring Tonality

Prior to every scheduled FOMC meeting, the staff puts together its forecast for the U.S. economy in a document called the *Greenbook* (now the *Tealbook*). Greenbook forecasts were published monthly up 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. In August 1974, the *Greenbook* was reorganized into Greenbook Part 1, the summary and outlook, which outlined the forecast, and Greenbook Part 2, which described recent developments. For most of our sample, text analysis is based on the text of Greenbook Part 1. Prior to August 1974, Greenbook was published as a single volume document which contained both the recent domestic developments and outlook for domestic economic activity as well as international developments; for these early observations we extract the 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 with Tealbook A, which consolidated the content in 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.² To derive our dictionary, we adopt the initial classification of positive and negative words in the widely used Harvard psycho-social dictionary³ 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

² For the list of positive and negative words, see the data appendix.

³ 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.

the words “demean” or “hedge” as negative. Positive words in our dictionary include terms like “enthusiasm,” “abundant,” “enhance,” and “successful.” On the other hand, 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.⁴

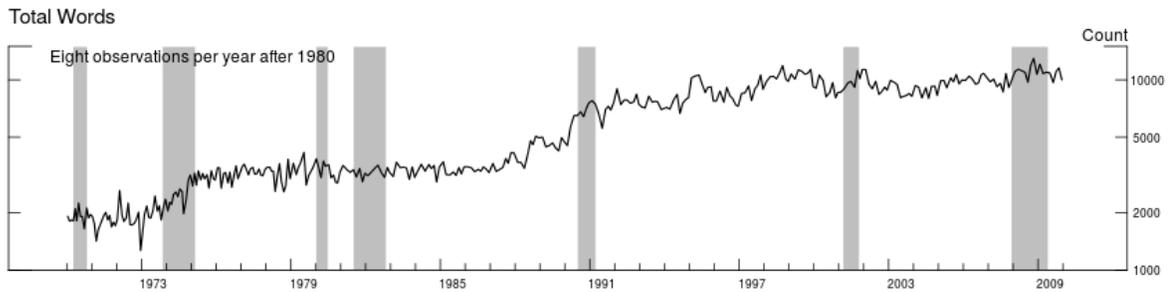
By using the whole document to quantify the overall degree of optimism, irrespective of how words are grouped, we are choosing not to use more elaborate methods of text analysis that would, for instance, attempt to identify double negatives or text specific to the economic indicators whose forecasts we evaluate.⁵ 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 forecasts on which we focus.

Figure 1 shows the time series of the total word counts from Greenbook Part I (or its equivalent prior to August 1974) through our 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.

⁴ 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.

⁵ As a 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.

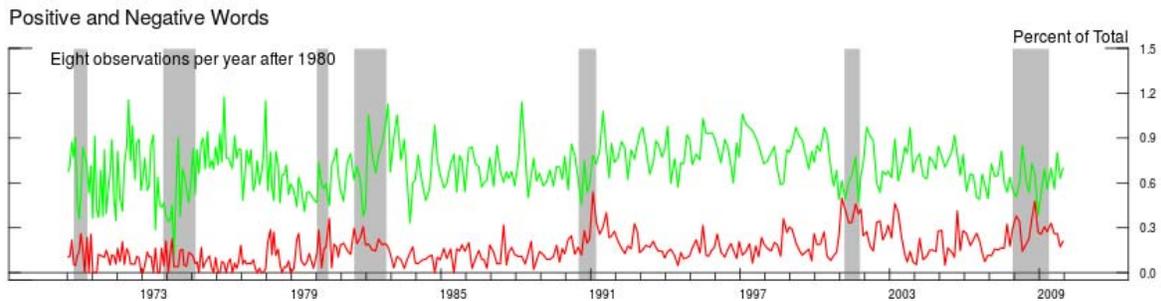
Figure 1: Total words in the Greenbook



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 August 1974 restructuring, the percentage of positive words per document appears to have been considerably more variable from one document to the next.

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, where 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.⁶ 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, except for very early in the sample, when there are not 40 previous comparable documents. For these observations, the corpus (of 40 documents) is defined to include past documents supplemented by nearby future documents.⁷

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. 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.

Figure 3 shows two side-by-side word clouds for the 50 most prominent positive words in Greenbooks during two periods, 1994-1998 and 2005-2009. The 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. The word cloud for positive words during 1994-1998 is slightly bigger than that for 2005-2009. Word choices between these two periods that are roughly ten years apart are similar, suggesting there is not a lot of language drift, whereby many words simply fall out of favor and are replaced by new ones. The most important positive word in both periods is “upward”, followed closely by “positive.” However, the word “favorable” is a more prominent word during 1994-1998, as is the word “moderation.”

⁶ 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.

⁷ 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.

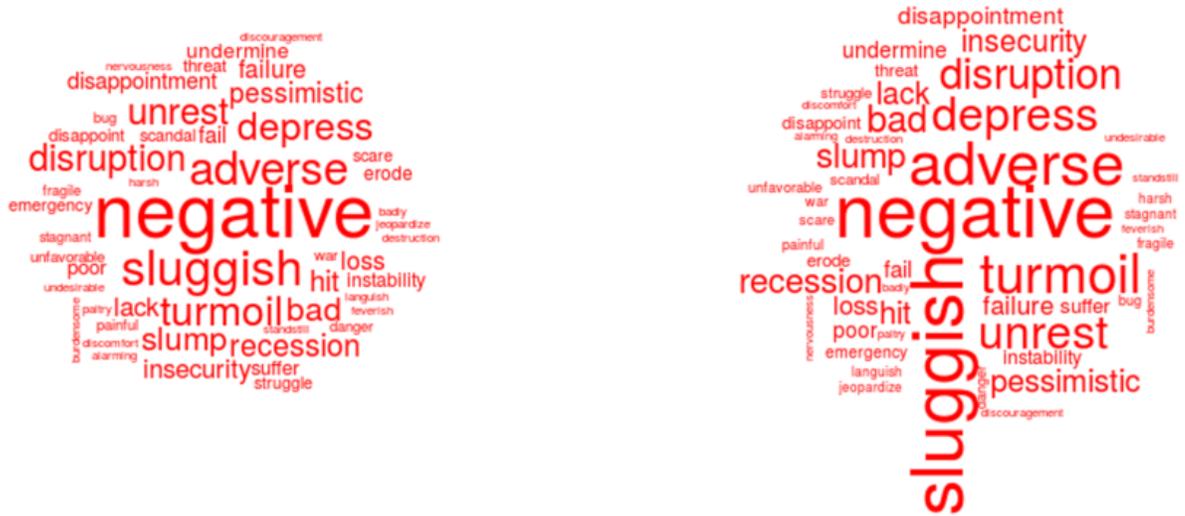
Figure 3: Word cloud for fifty most positive words in the Greenbook.



Note: The word cloud on the plot on left side shows fifty most positive words used in the Greenbook during the period Jan 1994 and Dec 1998. The word cloud on the right side shows fifty most positive words during the period Jan 2005 and 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 4 shows two side-by-side word clouds for the 50 most prominent negative words in Greenbooks during the same two periods. The word cloud for negative words during 1994-1998 is smaller than that in 2005-2009. The most prominent negative word in both samples is “negative”, followed by “sluggish.” However, negative words are simply more prominent in the latter sample as indicated by the larger word sizes in the right-hand-side word cloud. For example, the word “adverse” is somewhat more prominent in 2005-2009 period. Similarly the word “recession” is much smaller in the 1994-1998 period than it is in the 2005-2009 period, perhaps not surprising since the later period includes the “Great Recession.”

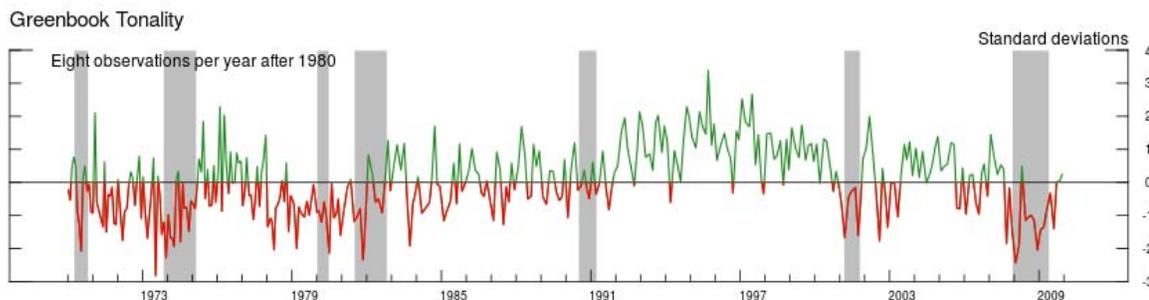
Figure 4: Word cloud for fifty most negative words in the Greenbook.



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

Figure 5 shows the Tonicity index plotted over the full sample period, with positive levels indicated in green and negative levels indicated in red. As one might expect, Tonicity 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 Tonicity are observations in the year leading up to and during the Great recession, and also during the 1974-75 recession. The most noticeable run of highly positive readings was during the mid-1990s. Despite these cyclical tendencies, Tonicity 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 5: Greenbook Tonality plotted over time

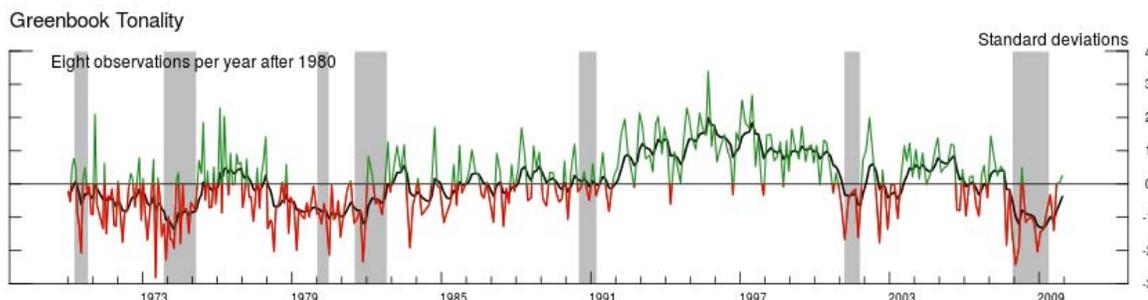


Note: Shaded regions represent NBER-dated recessions. Tonality is standardized to have a zero mean and a standard deviation equal to one. Tonality is shown in green when it is positive and in red when it is negative. Prior to 1981, Greenbooks were produced nearly every month, thereafter the frequency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

Considering the possibility that high-frequency movements could reflect more noise than signal, 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.⁸ 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 shocks that tend to reverse. We define deviations of Tonality from Trend Tonality as “Tonality Shocks.” **Figure 6** 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.

⁸ 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.

Figure 6: 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 Tonality is the black line overlayed on Tonality and tracks movements in Tonality.

B. Relation of Tonality to the Current Greenbook Point Forecasts

To explore links between the forecast text sentiment and the Fed staff quantitative forecast, we first examine simple correlations between Tonality and forecasts for three key economic performance variables: inflation, unemployment, and GDP growth. The first two are the two components of the Fed’s “dual mandate.” The third, GDP growth, is perhaps the most frequently cited summary statistic of economic performance. For each economic variable, we consider three metrics. The first is current economic conditions, specifically, forecasts of inflation, GDP growth, and the unemployment rate over the current quarter. Second, we construct gauges for the outlook four quarters ahead: cumulative inflation and GDP growth over the next four quarters, and the four-quarter change in the unemployment rate. Since the sentiment embedded in the narrative may be influenced both by the state of the outlook as well as by the direction of recent revisions, we also construct forecast revisions, relative to the previous Greenbook, for each outlook measure.⁹

The correlations between each of these forecast metrics and text Tonality—both raw Tonality and Trend Tonality—are shown in Table 1. The signs of the correlations between our

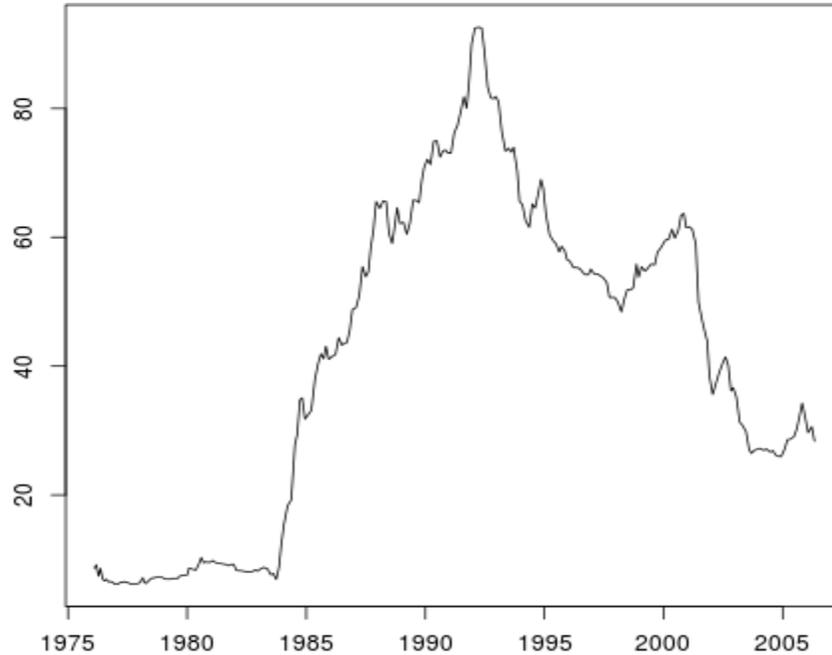
⁹ Revisions are measured as changes to the outlook only 3 quarters out. For most observations, constructing revisions to the 4-quarter outlook would require having the lagged value of the 5-quarter outlook, which is frequently unavailable.

measure of sentiment and our metrics of expected economic performance are consistent with intuition: negative for measures of expected inflation and unemployment but positive for measures of expected GDP growth. For all three variables, the four-quarter outlook is more strongly correlated with Trend Tonality than with raw Tonality. In contrast, for both GDP and unemployment, *revisions* to the outlook exhibit somewhat greater magnitude correlations with Tonality than with Trend Tonality. This is consistent with the conjecture that some of the innovations to Tonality might reflect temporary influences in the forecast.

We next explore the marginal contributions the economic forecast metrics have for “explaining” Tonality in a multivariate regression context (Table 3). To help keep this analysis tractable, we focus on the four-quarter outlooks for inflation and unemployment, as well as their revisions. We omit GDP outlook and its revisions because, not surprisingly, these are very highly (negatively) correlated with the respective measures of unemployment outlook, as shown in Table 2.

For the full sample (1972 – 2009), shown in the first column, we find that the Inflation Outlook, the Unemployment Outlook and the Unemployment Outlook revision all have marginal explanatory power for Tonality, with negative coefficients as intuition would suggest. All told, the quantitative metrics account for about 20 percent of the variation of Tonality over the full sample. To determine whether these relationships are hold similarly throughout the period of study, we use the Bai and Perron (2003) test to look for structural breaks in the econometric relationship between Tonality and the Fed’s forecast and forecast revisions. The test finds strong statistical evidence for a single break in the relationship, which is estimated to have occurred in January 1992. Indeed, a plot of the F-test values for all possible (single) breaks (Figure 7) suggests that the results regarding the timing of the break in the relationship between Tonality and economic forecast variables is fairly is quite definitive.

Figure 7: Chow tests for structural break in Tonality



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

The second and third columns show the Tonality regression estimates for the early (1972-1991) and late (1992-2009) sub-periods, respectively.¹⁰ Easily the most dramatic disparity between the factors driving text Tonality in the two sub-periods is a change in the sign on the Inflation Outlook. Prior to 1992, Inflation Outlook has a highly significant negative marginal effect on Tonality, whereas it has a significant positive marginal effect on Tonality in the later period. Although the positive effect of Inflation Outlook on Tonality post-1991 may seem 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 embedded in Fed forecast documents is connected to the so-called “Great Moderation.” While researchers commonly date the latter as occurring in the mid-1980s, that change might have been fully recognized and reflected in forecasts or forecast sentiment with some delay.

¹⁰ 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.

Rounding out the findings from the sample split, *revisions* to the inflation outlook do have a negative effect on Tonality during the later period. Another contrast in the regression results in these two periods is that, the negative effect on Tonality from the Unemployment Outlook and its revisions (found in the full sample) is quite strong in the later period but statistically insignificant in the early period. All told, these economic forecast variables explain four times the fraction of variation in the later period (adjusted R-squared of 44%) as compared to the early period (adjusted R-squared of 12%).

The last two columns show the multivariate relationship between the smoothed measure of sentiment, Trend Tonality, and the point forecasts for economic variables, using the same sample break as in the previous two columns. Consistent with the conjecture that Trend Tonality could be a less noisy measure of sentiment, the regression R-squared statistics for both subsamples rise markedly, to around 50% and 60% for the early and late periods respectively. Otherwise, results are mostly qualitatively similar to the Tonality regressions. In particular, the dichotomy in the inflation coefficient remains. One difference is that the Unemployment Outlook has a significant negative marginal effect on Trend Tonality in both the early and late sub-periods.

III. Greenbook Tonality as an Economic Indicator

Having established a strong contemporaneous connection between Tonality and expectations of key economic performance measures, our analysis turns to a central question of interest: does Tonality have predictive power for those 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 forecasted in Greenbook. We measure performance over 3 different horizons, the next quarter, two quarters ahead, and four quarters ahead. In each case, the dependent variable is the realized cumulative performance for the variable in question, and the explanatory variables are Tonality and the same-horizon Greenbook point forecast. In light of the structural change in how

Tonality of Greenbook text relates to inflation and GDP growth, regressions are estimated separately on the early and late subsamples.

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 examines 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. Following on this, “forecast efficiency” tests then examine whether adding other information variables to the 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 making the 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. The specifications would be identical if the coefficient on the Forecast_{t,t+h} was restricted to unity.

We also incorporate other information available at time t when the forecast is produced, which in previous studies have been found to improve upon the forecast or help to predict forecast error. One of these is the revision to the forecast from the previous Greenbook, which is motivated by findings of “information rigidities” by Coibion and Gorodnichenko (2012), who document that forecasts by the Survey of Professional Forecasts tend to be only partially adjusted toward their mean-square-error minimizing value. We also control for recent stock market returns in the regressions predicting economic variables (Stock and Waston 2003) and interest rate term spreads in regressions predicting the Fed funds rate (Rudebusch and Williams 2009).

A. Predicting GDP

Baseline regressions that examine the predictive content of Tonality for future GDP growth are shown for the early and late sample periods in Table 4, panel A and panel B, respectively. The dependent variable in the first column is realized 1-quarter-ahead GDP

growth. In the second and third columns it is realized cumulative GDP growth 2 and 4 quarters ahead, respectively. The first set of three regressions examines the predictive content of the GDP growth forecast by itself. Tonality is added in the second set of regressions, and in the third set Tonality is broken down into its trend and shock components. Standard errors shown below coefficient estimates 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)).

For the early subsample (the pre-1992 sample), coefficient estimates on the staff point forecast range from 0.99 for the 1 quarter-ahead forecast to 0.83 for the 4-quarter forecast, none of which is significantly different from 1.0. The R-squared statistics range from 0.64 for 1-quarter-ahead GDP to 0.48 for 4-quarter growth. Tonality is added to regressions shown in columns 4-6; the estimated coefficient is positive at all horizons but is only statistically significant in the four-quarter forecast regression, where the adjusted R-squared is marginally boosted from 0.48 to 0.49. When we break out the two components of Tonality in the last three columns, for the two longer horizons, the coefficient on Trend Tonality is substantially larger than that on the Tonality shock component, but neither is statistically significant.

Results are more interesting for the late period (Table 4 Panel B). For the initial benchmark regressions on forecast alone, the adjusted R-squared at each horizon is somewhat lower than in the early sample. Even so, the residual standard error is also quite a bit smaller in the later-period regression. Together, these statistics could reflect a reduction in predictable variation that came with the dampening of the business cycle, or “great moderation”. In columns 4-6, we find that Tonality has a significant positive coefficient for all three horizons, while the R-squared statistics rise compared to benchmark regressions, dramatically so at the 4-quarter horizon, from 0.24 to 0.34. At this horizon, the coefficient estimate implies that a one-standard deviation increase in Tonality boosts expected GDP growth by 67 basis points. What is more, out-of-sample R-squared statistics indicate that the benefit from incorporating Tonality largely carries over to a real-time exercise.

When we split Tonality into its trend and shock components, shown in columns 7-9, we find that Trend Tonality is the important component for predicting GDP growth, particularly for the 4-quarter-ahead forecast. Although coefficients on Trend Tonality (and Tonality Shock) are

similar to those in the early period, here they are statistically significant at the 2-quarter and 4-quarter horizon. Also, the adjusted R-squared for the four-quarter horizon is now 0.40, compared to the values of 0.24 and 0.34 in the baseline regression and the (undecomposed) Tonality regression. Looking over the full set of post-1991 sample regressions, another telling 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 fade the Greenbook point forecast somewhat, while putting some weight on the tone of the narrative in Greenbook, as quantified by Tonality.

To test the robustness of these results and their interpretation, we add to the regressions two control variables discussed earlier—the revision to the GDP forecast and stock market appreciation since the last FOMC meeting. Results are shown in Table 5. We do find some evidence that, in the late period, the recent forecast revision had some marginal predictive power for GDP growth 1 quarter ahead. This suggests that the GDP forecast was somewhat “sticky” in the sense of Coibion and Gorodnichenko (2012). In the later period, we see coefficient on GDP forecast revision decrease further and is only informative in predicting GDP one quarter ahead, suggesting a decline in forecast “stickiness.” In addition, regressions for both the early and later sample indicate that recent stock returns also have some positive marginal predictive power for GDP growth. At the same time, the coefficient estimates on Trend Tonality do not change materially when these controls are added.

B. Predicting Unemployment

Results from estimating the analogous regressions for the forecasted change in unemployment rate are shown in Table 6, Panel A and Panel B. Overall, findings regarding the predictive effects of Tonality are quite similar to those for GDP. In the early period, Trend Tonality has marginal predictive power for the 4-quarter-ahead change in unemployment (last column), where the adjusted R-squared rises to 0.63 compared to 0.61 in the baseline specification that conditions only on the point forecast. In the later subsample, Tonality, and particularly Trend Tonality, boost predictive power for all three horizons. The contribution at the 4-quarter horizon is quite substantial, with the R-squared rising from 0.52 in the benchmark regression to 0.57 with Tonality added and to 0.61 with Trend Tonality. At the same time, the coefficient on the point forecast for unemployment drops from an oversized 1.58 in the baseline

regression (3rd column) to 0.98 in the specification that includes Trend Tonality (last column). A one-standard deviation (0.72 units) increase in Trend Tonality reduces expected unemployment by about ½ percentage point. Echoing results for GDP, adding Tonality and Trend Tonality boosts the out-of-sample R2 statistics for unemployment forecast regressions in the later sample.

Adding the control variables (Table 7) has essentially no effect on the Tonality coefficients. The forecast revision has no marginal predictive power for realized unemployment. Recent stock returns have some marginal predictive power, but the estimated effects of Trend Tonality are robust to inclusion of these controls.

C. Predicting Inflation

Results from estimating the regressions for (cumulative) inflation forecast errors are shown in Table 8. Unlike our findings for GDP and Unemployment, the early (Panel A) and late (Panel B) results look qualitatively quite different, in two ways. First, Tonality has negative coefficients in the early period and positive coefficients in the late period. This echoes our findings in Table 3, which provided the rationale for the sample split. It is also consistent with the interpretation that higher inflation was a factor of independent concern prior to 1992, but not after then. Although not shown, results from regressions with controls added do not alter these conclusions.

Though somewhat tangential to focus of this paper, it is also interesting to note the very small coefficient estimates on the Staff Forecast in the 2-quarter and 4-quarter forecast horizons in the later period. For instance, the coefficient estimate of 0.06 on the four-quarter forecast (Panel B, column 3) and the negative R-squared implies that that forecast itself has no predictive power for actual inflation over the four quarters ahead. Together with the large positive intercept, these estimates suggests that forecast errors would be significantly reduced for the longer horizon forecasts if the forecast had called for a constant inflation rate equal to the expected average rate. This echoes findings by Atkeson and Ohanian (2001) and Stock and Watson (2007) that inflation has become harder to forecast since the mid-1980s.

D. Greenbook Tonality and Blue Chip Forecasts

So far, our findings indicate that Tonality of the Greenbook narrative has predictive value for GDP growth and unemployment, conditional on the Greenbook forecast. One question this raises is whether there is some built-in, perhaps even conscious, complementarity between the point forecast and the narrative. For instance, are there biases in the Greenbook point forecasts induced by some complementary communication built into the forecast narrative? While we cannot test this directly, it is possible to examine whether the predictive content of Tonality holds up when we instead condition on economic forecasts outside the Fed. If so, this would suggest that, even by itself, the information content measured by Tonality would be valuable, not only as a complement to the Greenbook point forecast but also as an adjunct to other economic or financial market participants.

We use the consensus Blue Chip Financial Forecasts to conduct such an exercise. Of course, doing so requires contending with an imperfect match in the timing of when the Blue Chip and the Greenbook forecasts are published. We match as follows: When the Greenbook forecast is published on or before the 15th of the month, then that Greenbook (its Tonality) is married with the most recent (previous-month) Blue Chip forecast; otherwise, it is married with the upcoming (end-of-current-month) Blue Chip forecast. While the analysis was conducted using all three forecast horizons and on both sub-periods, for brevity, we focus on the 2-quarter and 4-quarter forecast horizons in the post-1991 sample only, given that the Blue Chip Financial Forecasts were only available beginning in 1984.

The dependent variables in the first two columns of Table 9 are realized 2- and 4-quarter GDP growth, while explanatory variables include the Blue Chip GDP growth forecast for the same horizon, and the components of Greenbook Tonality; regressions are estimated with the additional controls, though we get very similar results without controls. The second and third pairs of columns show analogous forecast regressions for the change in unemployment and the inflation rate. In each case, the results are remarkably similar to the analogous regressions that conditioned on the Greenbook forecast. When conditioning on the Blue Chip forecast, Trend Tonality is again a statistically positive predictor of future GDP growth, a strong negative predictor for future unemployment, and a positive predictor for future inflation. These results suggest that the predictive content in the Tonality of the Greenbook forecast narrative does not seem to owe to a unique complementarity between Greenbook point forecasts and Tonality.

IV. Tonality as a Predictor of Monetary Policy

Given that Tonality is helpful for predicting economic performance up to four quarters ahead, a reasonable corollary hypothesis would seem to be that Tonality has predictive power for monetary policy over a similar horizon. In particular, in the post-1991 period, higher Tonality tends to signal stronger future economic activity relative to economic point forecasts, whether those forecasts are by Fed staff or the private sector. As a consequence, one might expect higher Tonality to predicate higher policy rates, and perhaps even higher-than-forecast policy rates.¹¹

The logic of the hypothesis that Tonality would predict surprises in the Fed funds rate seems straightforward, but there are potential differences between the analysis that conditions on funds rate forecasts in Greenbook versus those by private (Blue Chip) forecasters. To the extent that Blue Chip consensus forecasts of interest rate policy are connected to Blue Chip consensus forecasts for the economy (such as through a perceived Taylor rule), then positive economic surprises presaged by Tonality should, in turn, presage positive surprises in the path of policy rates. The logic for such a connection is somewhat murkier in the case of funds rate forecast surprises relative to Greenbook, particularly in light of arguments by Reifschneider and Tulip (2017). They 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 suggests that Greenbook funds rate forecasts will tend to appear timid relative to a prescriptive forecast, providing an additional rationale for why Tonality might help to predict surprises in the funds rate relative to the Greenbook forecast.

Our baseline funds rate forecast regressions that condition on Greenbook point forecasts for the funds rate are shown in Table 10. In these regressions, the realized change in the funds rate is regressed on the Greenbook forecast of the change in funds rate (over the same horizon),. As shown in the first three columns, which include solely the Staff Forecast, coefficients on the point forecast are significantly larger than 1.0, dramatically so at longer horizons. This seems consistent with the claim by Reifschneider and Tulip (2017) that Greenbook funds rate forecasts

¹¹ We focus only on the post-1990 period (the “late sample”) for two reasons: first, this is where Tonality was found to have robust predictive power for economic variables, and second, because Greenbook and Blue Chip forecasts of the Fed funds rate are only available beginning in 1983 and 1984, respectively.

tend to be timid. As hypothesized, when Tonality is added to regressions in the subsequent three columns, its coefficients are positive and significant at all three horizons; thus, Tonality helps forecast the funds rate. Improvement to the regression fit is modest, however, with the R-squared for the 4-quarter-ahead funds rate rising from 0.47 to 0.49. Finally, in contrast to our results for GDP and Unemployment, splitting Tonality into its two components (the final three columns) does not improve regression fit, as the coefficients on the Trend and Shock components are not statistically distinguishable.

Analogous tests that condition on Blue Chip funds rate forecasts are shown in Table 11, with results that are remarkably similar. For all three horizons, the coefficient on the forecast in the basic specification is again substantially and significantly higher than 1.0. Interestingly, though, Blue Chip forecasts appear to be somewhat more correlated with realized changes in the fund rate compared to the Greenbook Fed funds forecasts, as indicated by the somewhat higher R-squared statistics. Nonetheless, similar to the regressions that condition on the Greenbook forecast, when Tonality is added, it has statistically significant positive coefficients. Also similar, the decomposition of Tonality into Trend and Shock does not boost predictive power except in the 4-quarter horizon regression.

Table 12 shows similar tests but with additional variables that we would expect to help forecast changes in the funds rate. As with the analysis of economic forecasts, we control for forecast rigidity using the revision in the funds rate forecast relative to the previous Greenbook forecast (or, in the case of Blue Chip, relative to the Blue Chip forecast nearest to the previous Greenbook). We also add *Term Spread*, a measure of the market expectation for the change in the funds rate over a comparable horizon. This allows us to test whether economists' forecasts of the funds rate efficiently reflect market expectations and, at the same time, will help assess the extent to which Tonality's predictive power is related to market expectations. Following Gürkaynak, Sack and Swanson (2007), we gauge market expectations as the spread between Fed funds futures contract rates and the current funds rate, using the futures contract maturity that best approximates the forecast horizon.¹²

¹² For the 1- and 2-quarter-ahead funds rate forecast, market expectations should be well approximated using the futures contract maturing 3 and 6 months ahead, respectively. For the 4-quarter-ahead forecast, we again use the 6-

As shown in the first three columns, the coefficients on the Greenbook funds rate forecast drop dramatically relative to the regressions that did not include futures rates, while the coefficients on the Term Spread in each case are large and statistically significant. These results unequivocally indicate that the Greenbook funds rate forecast does not reflect the market's information about the likely path for the funds rate. The coefficients on tonality drop and remains a significant predictor of the future funds rate only at the 1-quarter horizon, suggesting that Tonality contains limited information with respect to the funds rate not reflected in the futures rates.

The remaining columns show the results when regressions are conditioned on the Blue Chip forecast. Here again, we find that the Fed funds futures rates help predict the realized funds rate at all three horizons, though with somewhat smaller coefficients. These results suggests that Blue Chip forecasts also do not reflect all the information about subsequent policy signaled by markets. Also similar to the Greenbook fed funds rate forecasts, including the Term Spread reduces the marginal predictive power of Tonality; however, Tonality still remains for two of the three horizons, providing at least some support for the conjecture that Greenbook Tonality has predictive power for monetary policy as well as for real economic variables, which might not be fully reflected in financial market prices.

V. Tonality and Future Stock Returns

The evidence from the previous two sections indicating that economists' forecasts do not contain all the information embedded in Greenbook Tonality begs the question: do asset market prices reflect all relevant information in Tonality? If not, 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. Indeed, given that we already have shown Tonality helps predict

month ahead futures rate, the latest-maturing reliably-traded contract. Alternatively, using the 4-quarter-ahead Eurodollar futures rate for the latter does not materially change the results.

some innovations to Fed funds rates, the implications for bond return forecasting seem potentially quite rich and deserving of careful attention, which we reserve for future study.

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 13 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 and Tonality shock. 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 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). In contrast, Tonality Shock has no predictive content, consistent with the results of our economic forecast regressions.

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 R² 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.¹³

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 as a proxy for investors' risk premium, which would have the odd implication that investors demand a higher risk premium when Greenbook sentiment is more positive. 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.)

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. Indeed, the staff forecast of current unemployment has a correlation of 99% with the analogous and publicly observable Blue Chip forecast, where the two overlap in our sample. 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

¹³ 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_i), 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.

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.

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. The last three columns show results when Current Unemployment is used as a predictor on its own. Here, its significance disappears in the longest-horizon predictions; and, while the in-sample R-squared statistics remains respectable, all the out-of-sample R-squared statistics turn negative. What is more, while not shown here, if we also include other standard predictors such as the dividend yield in these regression, Current Unemployment is no longer significant, while the predictive value of Trend Tonality remains robust.

VI. Deeper Look into Tonality Effects

This final section provides some additional color on the information content of our measure of sentiment in the Greenbook text by looking for evidence that might support the idea that the Tonality signals something about the balance of risks to the point forecast. The typical point forecast in the Greenbook, as well as in the Blue Chip survey, is often viewed as representing a modal forecast, rather than a mean forecast designed to minimize mean squared

errors. If so, then Tonality might reflect the relative importance of various non-modal outcomes, or even “tail” outcomes. For instance, it is well-known that quantitative economic forecasts during expansions rarely project recessions. Perhaps Tonality reflects the perceived risk of a recession.

To test this interpretation, we first examine whether Tonality helps forecast the incidence of recession, while controlling for the point forecasts for unemployment or GDP growth. In order to avoid over-fitting, which is a risk given the very small number of recessions in each sub-period, we conduct this test using the full sample. We estimate probit models for recession 3, 6, and 12 months ahead, using the NBER definition of recessions for the U.S. economy from 1973-2009, results from which are shown in Table 14.

In the basic specification for each horizon, we include the staff forecast for unemployment change and GDP growth rate over the matching horizon (1, 2, or 4 quarters) and an indicator variable to control for whether the economy was already in recession 2 months before the forecast was made. In short, we find that Trend Tonality is a significant predictor of recessions, with a p-value below 1 percent at the first two horizons and below 5 percent at the latest. Indeed, with Trend Tonality in the regression, the point forecasts for unemployment and GDP have very little, if any, marginal predictive content beyond the first horizon. Even when we add two well-established financial market based predictors, the term spread (10 year minus 1 year Treasury yield) and the GZ (credit) spread (Gilchrist and Zakrajšek 2012), Tonality’s predictive power remains statistically significant, bolstering our inference that Tonality provides a valuable signal of downside risk.

A different 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 can estimate quantile regressions in which the dependent variable is the realized forecast error in Greenbook—for either the unemployment rate or GDP growth—and Tonality is the explanatory variable. The first two columns in Table 15 show the results for the unemployment rate forecast error, regressed on either Tonality or Trend Tonality, over the full sample. (Results are qualitatively similar for the post-1991 sample.) As shown, compared to its value at the median, the negative coefficient is substantially larger in magnitude at the 75th percentile and especially

at the 90th percentile statistics indicate that the latter difference is highly significant.¹⁴ These results imply that Tonality provides a particularly strong signal when unemployment comes in higher than forecast, that is, when the economic outcome is worse than expected.

Results for GDP forecast errors shown in third and fourth columns are analogous, with tonality having a larger coefficient at the lower quantiles—again, the bad news end of the spectrum. In the case of GDP, the difference in coefficient between the 50th and 10th percentiles is not statistically significant. However, the qualitative progression of the coefficient as one moves from higher to lower quantiles is striking, and when the test is extended to compare the 10th and the 75th percentiles, we do find a significant difference. Finally, the last two columns examine quantile regressions for the staff’s Fed Funds rate forecast errors. Here the qualitative pattern indicates that Tonality has a bigger effect at both tails, compared to at the median; but none of the differences are statistically significant.

Overall, the quantile regressions appear fairly supportive of the conjecture that a disproportionate amount of the information conveyed by Tonality (and Trend Tonality) is related to downside risks to the economy. These results thus beg the question: is the predictive value of Tonality for stock returns similarly most evident for returns closer to the lower tail of the distribution of returns? If so, this would provide support for our conjecture that Tonality’s predictive value for stock returns likely derives from the information Tonality contains about future economic performance that is not yet incorporated into market prices at the time the Greenbook was produced. Table 16 shows estimates from quantile (prediction) regressions for stock returns, conditioned on Trend Tonality. For all three horizons, the coefficients on Trend Tonality are indeed largest at the 10th percentile of returns, and decline monotonically as we move to higher quantiles. As shown at the bottom of the table, tests for differences in the coefficient on Trend Tonality from its value at the median show that, for both 3-month and 6-month returns, the coefficient on Tonality is significantly larger at the 10th quantile (compared to at the median), and significantly smaller at the 90th quantile.

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.

The final question we consider is whether there is different information in negative words than in positive words. For instance, Loughran and McDonald (2011) find that in annual reports of U.S. firms, negative words are more informative than positive words. In our construction, Tonality equals the amount of positive Tonality minus the amount of negative Tonality; thus, these components can easily be decoupled. For the exercise in Table 17, we estimate a decomposition for each of the two signed components of Tonality into trend and shock components, yielding four variables, Trend Positivity, Positivity Shock, Trend Negativity and Negativity Shock. We then estimate regressions predicting 2-, and 4-quarter GDP growth and Unemployment change and stock prices on the four components.

For realized GDP growth, the predictive information is almost exclusively coming from Trend Positivity. For the current quarter, the hypothesis that the positive and negative trend components have equal and opposite coefficients can be rejected with 10 percent confidence, and for two and four quarters ahead the hypothesis can be rejected at 5 and 1 percent confidence, respectively. A different picture emerges for the unemployment forecast, where we find that, at least for the longer two horizons, both the positive and negative component of Tonality each contribute with the expected sign. High Positivity predicts a lower unemployment rate relative to the staff forecast; and High Negativity predicts a higher unemployment rate. Furthermore, we cannot reject the hypothesis that Positivity and Negativity have equal and opposite signs for the two horizons. For stock price changes, we cannot reject the hypothesis that the two pieces have equal and opposite effects, but only Trend Positivity continues to show statistical significance at each horizon.

VII. Summary, Interpretation, and Conclusions

The predictive contribution of the Tonality of the Greenbook text for unemployment and GDP growth, when conditioning on the Greenbook forecast for those variables, suggests that an important element of economic forecasting is found 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 seems to reflect either stickiness in the forecast or information signaled by recent stock price movements.

The fact that a component of Tonality predicts monetary policy surprises in the fed funds futures indicates that Tonality conveys policy-relevant information. The finding that Tonality predicts equity prices over subsequent 3-, 6- and 12-month periods is notable but perhaps 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, this return predictability would not seem to reflect compensation for risk. Rather, these results suggest that equity prices do not contemporaneously impound all the information about the economy that has been aggregated into the forecast narrative.

The evidence presented in this paper argues for examining forecast effectiveness while including other information forecasters are relying along with the quantitative forecasts. Doing so will require preserving (and in some cases) obtaining the narrative accompanying the forecasts. Quantile regressions for forecast errors and recession probability regressions 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 of economic forecast is informative in 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.

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 Econometrics* 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 Gorodnichenko. 2012. "What can survey forecasts tell us about information rigidities?" *Journal of Political Economy* 120 (1): 116-159.
- 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.
- 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 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 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.
- Lettau, Martin, and Sydney Ludvigson. 2001. "Consumption, Aggregate Wealth, and Expected Stock Returns." *The Journal of Finance* 56 (3): 815-849.
- Loughran, Tim, 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.
- Newey, Whitney K., and Kenneth D. West. 1994. "Automatic lag selection in covariance matrix estimation." *The Review of Economic Studies* 61 (4): 631-653.
- 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.
- 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.

Data Appendix:

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. For the dictionary, we used the Harvard psycho-social dictionary as a base, 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
conductive	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	

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:

$$Act_Unemployment_change_i = Unemployment_i - Unemployment_{-1}$$

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:

$$Act_CPI_growth_i = 100 * ((Act_CPI_i / Act_CPI_0) - 1)$$

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 published by Wolters Kluwer Legal and Regulatory Solutions U.S., 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:

$$Act_FedFunds_change_i = Act_FedFunds_i - Act_FedFunds_{-1}$$

Staff Forecasts

Staff projections for the Fed funds rate are from the financial assumptions dataset maintained by the Philadelphia Fed from 1992-2009.

$Staff_FedFunds_{-1}$ is equal to the staff's forecast for the previous quarter. $Staff_FedFunds_i$ is equal to the staff's forecast i quarters into the future. We define the change in the Fed funds rate as follows:

$$Staff_FedFunds_change_i = Staff_FedFunds_i - Staff_FedFunds_{-1}$$

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.

$$BC_FedFunds_change_i = BC_FedFunds_i - BC_FedFunds_{i-1}$$

Term Spreads

The term spreads variables are calculated following (Gürkaynak, Sack and Swanson 2007) as the spread between Fed funds futures contract rates and the current Fed funds rate. The Fed funds future rate is chosen to align most closely with the forecast period given the availability of the contract. In the case of 1-quarter ahead, the 3-month contract is used and for the 2- and 4- quarter ahead forecasts, the 6-month contract is used. The spread is taken at the Greenbook release date.

$FedFutures_i$ is equal to the Fed funds futures contract aligned with the forecast i quarters into the future. $Curr_FedFunds$ is the current Fed funds rate. We define the term spread as follows:

$$Term_Spread_i = FedFutures_i - Curr_FedFunds$$

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 i th to the j th FOMC Date.

6. Recession Variables

Recessions are defined using the NBER's recession dates. All data uses the full sample of 1973-2009.

$Recession_i$ is a dummy equal to 1 if the United States is in a recession i months ahead

$Prev_Recession$ is a dummy equal to 1 if the United States was in a recession 2 months ago.

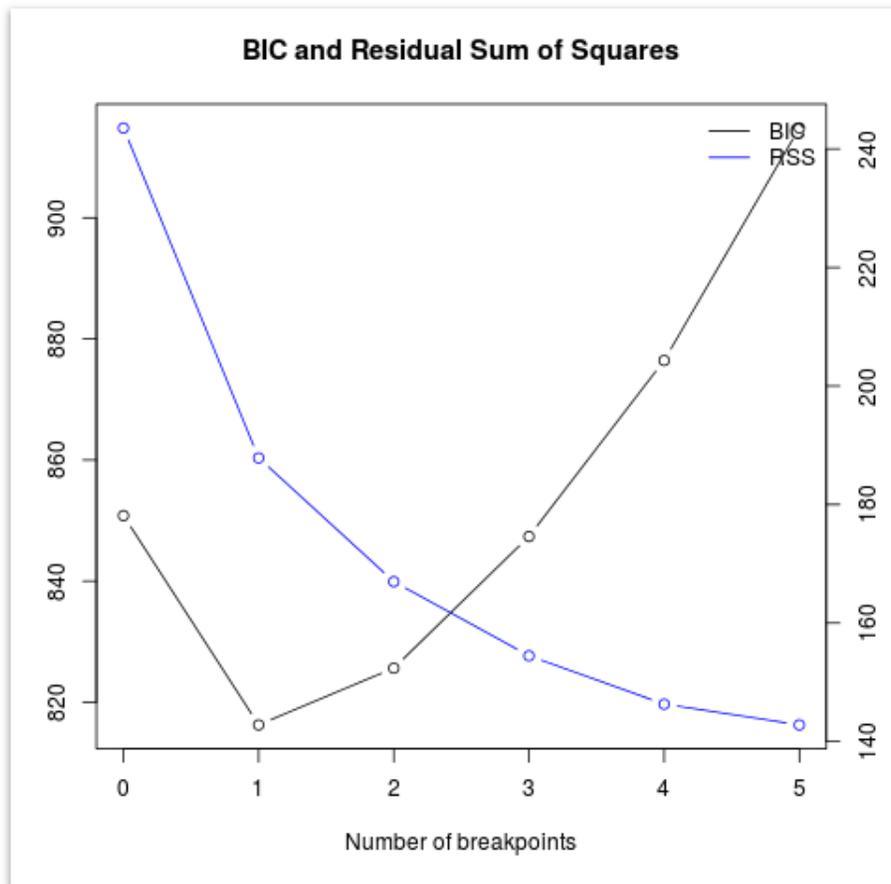
$Term_Spread_Recession$ is equal to the difference between the 10-year treasury yield and the 1-year treasury yield.

GZ_Spread is defined in Gilchrist and Zakrajsek (2012).

Appendix: 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 December, 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 A1: 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 December 1991.

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

	Tonality	Trend Tonality
Current GDP Growth	0.17***	0.29***
GDP Outlook	0.22***	0.29***
GDP Outlook Revision	0.26***	0.20***
Current Unemployment	-0.07	-0.24***
Unemployment Outlook	-0.33***	-0.46***
Unemployment Outlook Revision	-0.27***	-0.25***
Current Inflation	-0.32***	-0.43***
Inflation Outlook	-0.33***	-0.49***
Inflation Outlook Revision	-0.10*	-0.13**

Notes: Current GDP growth is the GDP growth rate for the current quarter as expected by the staff forecast in the Greenbook, GDP outlook is the cumulative 4-quarter GDP growth between the next quarter and 4-quarters later. GDP outlook revision is the revision to the GDP outlook from previous Greenbook to the current Greenbook. Unemployment and Inflation outlook and revision variables are similarly defined with respect to the unemployment rate and inflation rate. *p<0.1; **p<0.05; ***p<0.01

Table 2: Pearson Correlations among Greenbook forecast variables

	Current GDP	GDP Outlook	GDP Rev	Current Unemp	Unemp Outlook	Unemp Rev	Current Infl	Infl Outlook
Current GDP								
GDP Outlook	0.79***							
GDP Rev	0.28***	0.30***						
Current Unemp	-0.08	0.11*	0.09					
Unemp Outlook	-0.82***	-0.86***	-0.33***	-0.11**				
Unemp Revision	-0.39***	-0.31***	-0.70***	0.07	0.42***			
Current Infl	-0.15***	-0.28***	-0.02	0.19***	0.33***	-0.04		
Infl Outlook	-0.15***	-0.27***	-0.01	0.36***	0.28***	-0.03	0.84***	
Infl Rev	0.06	-0.07	0.01	-0.06	0.09	-0.08	0.29***	0.29***

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

Table 3: Greenbook Forecast Factors in Text Tonality

	Tonality			Trend Tonality	
	Full sample	Up to 1991-12-11	Post 1991-12-11	Up to 1991-12-11	Post 1991-12-11
Inflation Outlook	-0.128*** (0.023)	-0.117*** (0.040)	0.466*** (0.104)	-0.096*** (0.015)	0.482*** (0.063)
Inflation Rev	-0.045 (0.180)	-0.087 (0.212)	-0.751*** (0.279)	0.015 (0.080)	-0.677*** (0.170)
Unemployment Outlook	-0.183** (0.073)	-0.065 (0.081)	-0.749*** (0.149)	-0.157*** (0.031)	-0.762*** (0.091)
Unemployment Outlook Rev	-0.923*** (0.226)	-0.344 (0.248)	-1.137*** (0.363)	-0.053 (0.094)	-0.254 (0.221)
Intercept	0.675*** (0.109)	0.473** (0.237)	-0.449* (0.251)	0.345*** (0.090)	-0.494*** (0.153)
Observations	317	173	144	173	144
Adjusted R ²	0.206	0.116	0.441	0.504	0.597
Residual Std. Error	0.915	0.811	0.806	0.307	0.490
F Statistic	21.473***	6.664***	29.162***	44.748***	53.955***

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

Table 4: Regressions Predicting Cumulative GDP Growth

Quarters Ahead	1	2	4	1	2	4	1	2	4
<i>Panel A. 1970-1991</i>									
Staff Forecast	0.99*** (0.10)	0.98*** (0.13)	0.83*** (0.14)	0.98*** (0.10)	0.97*** (0.13)	0.79*** (0.13)	0.98*** (0.09)	0.94*** (0.11)	0.68*** (0.17)
Tonality				0.16 (0.11)	0.17 (0.16)	0.48* (0.28)			
Trend Tonality							0.16 (0.32)	0.45 (0.65)	1.61 (1.26)
Tonality Shock							0.17 (0.11)	0.05 (0.17)	0.06 (0.30)
Intercept	-0.11 (0.24)	-0.20 (0.46)	0.11 (0.77)	-0.04 (0.22)	-0.13 (0.42)	0.38 (0.72)	-0.05 (0.18)	0.02 (0.35)	1.04 (0.90)
P(Forecast = 1)	0.96	0.89	0.22	0.83	0.80	0.11	0.81	0.60	0.06
Observations	215	214	174	215	214	174	215	214	174
Adjusted R ²	0.64	0.59	0.48	0.65	0.59	0.49	0.65	0.59	0.51
Residual Std. Error	1.11	1.60	2.29	1.11	1.59	2.26	1.11	1.59	2.22
Out-of-sample R ²	0.44	0.43	0.44	0.43	0.41	0.43	0.40	0.38	0.43
<i>Panel B. 1992-2009</i>									
Staff Forecast	0.92*** (0.16)	0.89*** (0.18)	0.73*** (0.25)	0.83*** (0.15)	0.77*** (0.15)	0.57*** (0.19)	0.80*** (0.15)	0.72*** (0.14)	0.53*** (0.19)
Tonality				0.14* (0.08)	0.29** (0.12)	0.67** (0.27)			
Trend Tonality							0.22 (0.14)	0.48* (0.25)	1.13** (0.47)
Tonality Shock							0.07 (0.08)	0.06 (0.12)	0.06 (0.19)
Intercept	0.31 (0.25)	0.45 (0.44)	1.11 (1.10)	0.35 (0.24)	0.55 (0.39)	1.32 (0.84)	0.35 (0.24)	0.55 (0.37)	1.22 (0.80)
P(Forecast = 1)	0.62	0.55	0.27	0.27	0.13	0.02	0.18	0.05	0.01
Observations	144	144	144	144	144	144	144	144	144
Adjusted R ²	0.50	0.43	0.24	0.52	0.47	0.34	0.52	0.48	0.40
Residual Std. Error	0.73	1.05	1.80	0.72	1.02	1.68	0.72	1.00	1.61
Out-of-sample R ²	0.45	0.37	0.21	0.50	0.44	0.34	0.50	0.47	0.39

Notes: Estimates from the regression of 1-,2- and 4- quarter cumulative GDP growth on Fed Staff forecast, and Tonality (or Trend and Shock components of Tonality). Cumulative growth rate in GDP is measured from the current quarter ($k=0$). Panel A shows the estimates between 1970 and December 1991; Panel B shows the estimates after January 1992 to December 2009. Trend and Shock components of Tonality are derived by constructing an exponentially weighted moving average of Tonality. 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 R² 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 Cumulative GDP Growth with Controls

Quarters Ahead	1970-1992			1992-2009		
	1	2	4	1	2	4
Staff Forecast	0.95*** (0.09)	0.90*** (0.11)	0.68*** (0.17)	0.71*** (0.12)	0.64*** (0.14)	0.50** (0.22)
Trend Tonality	0.15 (0.29)	0.41 (0.59)	1.25 (1.12)	0.15 (0.12)	0.41* (0.23)	1.06** (0.47)
Tonality Shock	0.11 (0.11)	-0.04 (0.18)	0.04 (0.32)	-0.11 (0.08)	-0.11 (0.10)	-0.07 (0.16)
Staff Revision	0.37 (0.26)	0.45 (0.32)	0.31 (0.30)	0.36** (0.17)	0.34 (0.32)	0.05 (0.49)
Recent Stock Return	0.03 (0.02)	0.07** (0.03)	0.06* (0.03)	0.05*** (0.01)	0.05** (0.02)	0.06*** (0.02)
Intercept	0.003 (0.19)	0.07 (0.36)	0.97 (0.91)	0.46** (0.19)	0.70* (0.38)	1.31 (0.94)
P(Forecast = 1)	0.54	0.40	0.06	0.02	0.01	0.02
Observations	214	211	171	144	144	144
Adjusted R ²	0.66	0.62	0.52	0.59	0.52	0.41
Residual Std. Error	1.09	1.53	2.13	0.67	0.96	1.59
Out-of-sample R ²	0.40	0.28	0.41	0.54	0.48	0.40

Notes: Estimates from the regression of 1-,2- and 4-quarter cumulative GDP growth on Fed Staff forecast, and Tonality (or Trend and Shock components of Tonality), revision to the forecast, and recent stock return. Recent stock return is the stock return from the prior Greenbook to the Greenbook. First three columns show estimates between 1970 to December 1991. The last three columns show estimates after January 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 R² 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 Unemployment Change

Quarters Ahead	1	2	4	1	2	4	1	2	4
<i>Panel A. 1970-1991</i>									
Staff Forecast	1.01*** (0.10)	1.05*** (0.12)	1.09*** (0.15)	1.00*** (0.10)	1.05*** (0.12)	1.05*** (0.15)	0.99*** (0.08)	1.00*** (0.11)	0.87*** (0.15)
Tonality				-0.05 (0.04)	-0.02 (0.05)	-0.14 (0.09)			
Trend Tonality							-0.07 (0.12)	-0.18 (0.20)	-0.77* (0.42)
Tonality Shock							-0.04 (0.04)	0.04 (0.05)	0.08 (0.11)
Intercept	-0.05 (0.04)	-0.07 (0.07)	-0.09 (0.14)	-0.07 (0.04)	-0.08 (0.08)	-0.12 (0.14)	-0.07 (0.05)	-0.12 (0.09)	-0.26 (0.17)
P(Forecast = 1)	0.92	0.69	0.58	1.00	0.71	0.73	0.93	0.99	0.40
Observations	215	214	174	215	214	174	215	214	174
Adjusted R ²	0.67	0.64	0.60	0.67	0.64	0.61	0.67	0.64	0.63
Residual Std. Error	0.45	0.62	0.87	0.45	0.63	0.87	0.45	0.62	0.84
Out-of-sample R ²	0.62	0.62	0.55	0.62	0.62	0.55	0.62	0.61	0.55
<i>Panel B. 1992-2009</i>									
Staff Forecast	1.19*** (0.12)	1.40*** (0.19)	1.58*** (0.27)	1.13*** (0.11)	1.30*** (0.15)	1.23*** (0.16)	0.99*** (0.12)	1.09*** (0.14)	0.98*** (0.18)
Tonality				-0.05 (0.03)	-0.08 (0.06)	-0.32** (0.15)			
Trend Tonality							-0.14** (0.07)	-0.27** (0.14)	-0.66** (0.31)
Tonality Shock							0.02 (0.04)	0.05 (0.05)	-0.03 (0.08)
Intercept	-0.08*** (0.03)	-0.10* (0.05)	-0.10 (0.14)	-0.05 (0.04)	-0.05 (0.08)	0.11 (0.21)	0.01 (0.06)	0.07 (0.12)	0.32 (0.29)
P(Forecast = 1)	0.12	0.04	0.04	0.25	0.05	0.15	0.95	0.51	0.91
Observations	144	144	144	144	144	144	144	144	144
Adjusted R ²	0.74	0.69	0.52	0.74	0.70	0.57	0.75	0.72	0.61
Residual Std. Error	0.29	0.45	0.85	0.29	0.44	0.80	0.28	0.43	0.76
Out-of-sample R ²	0.72	0.66	0.47	0.73	0.67	0.52	0.74	0.68	0.52

Notes: Notes: Estimates from the regression of 1-,2- and 4- quarter change in the unemployment rate on Fed Staff forecast of unemployment rate change, and Tonality (or Trend and Shock components of Tonality). The change in the unemployment rate is measured with respect to the current quarter estimate in the Greenbook. Panel A shows estimates between 1970 to December 1991; Panel B shows estimates after January 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 R² 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: Regressions Predicting Unemployment Change with Controls

Quarters Ahead	1970-1992			1992-2009		
	1	2	4	1	2	4
Staff Forecast	0.99*** (0.09)	1.01*** (0.10)	0.89*** (0.15)	0.92*** (0.13)	1.02*** (0.14)	0.92*** (0.20)
Trend Tonality	-0.06 (0.11)	-0.15 (0.18)	-0.67* (0.37)	-0.13** (0.06)	-0.25** (0.13)	-0.63** (0.30)
Tonality Shock	-0.03 (0.04)	0.06 (0.05)	0.10 (0.12)	0.07 (0.04)	0.11* (0.06)	0.05 (0.09)
Staff Revision	0.05 (0.14)	-0.02 (0.16)	-0.04 (0.23)	0.21 (0.18)	0.18 (0.24)	0.07 (0.29)
Recent Stock Return	-0.01 (0.01)	-0.03** (0.01)	-0.03** (0.01)	-0.01* (0.01)	-0.02* (0.01)	-0.03** (0.01)
Intercept	-0.06 (0.05)	-0.09 (0.09)	-0.22 (0.16)	0.03 (0.06)	0.09 (0.12)	0.33 (0.29)
P(Forecast = 1)	0.87	0.94	0.47	0.51	0.86	0.71
Observations	214	211	171	144	144	144
Adjusted R ²	0.67	0.65	0.64	0.77	0.74	0.62
Residual Std. Error	0.45	0.61	0.82	0.27	0.41	0.75
Out-of-sample R ²	0.62	0.58	0.55	0.75	0.70	0.55

Notes: Estimates from the regression of 1-,2- and 4- quarter change in the unemployment rate on Fed Staff forecast of unemployment rate change, and Tonality (or Trend and Shock components of Tonality), revision to the forecast, and recent stock return. Recent stock return is the stock return from the prior Greenbook to the Greenbook. First three columns show estimates between 1970 to December 1991. The last three columns show estimates after January 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 R² 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 8: Regressions Predicting Inflation

Quarters Ahead	1	2	4	1	2	4	1	2	4
<i>Panel A. 1970-1991</i>									
Staff Forecast	0.78*** (0.11)	0.80*** (0.10)	0.87*** (0.15)	0.74*** (0.11)	0.77*** (0.10)	0.81*** (0.16)	0.67*** (0.12)	0.66*** (0.12)	0.72*** (0.24)
Tonality				-0.11** (0.05)	-0.14 (0.11)	-0.39 (0.30)			
Trend Tonality							-0.33* (0.17)	-0.72* (0.44)	-1.10 (1.25)
Tonality Shock							-0.03 (0.06)	0.08 (0.10)	-0.16 (0.17)
Intercept	0.31** (0.15)	0.60** (0.30)	0.78 (0.96)	0.33** (0.15)	0.64** (0.30)	1.04 (1.05)	0.37** (0.15)	0.78** (0.33)	1.41 (1.37)
P(Forecast = 1)	0.05	0.06	0.40	0.02	0.03	0.25	0.00	0.01	0.25
Observations	215	214	174	215	214	174	215	214	174
Adjusted R ²	0.47	0.46	0.47	0.48	0.47	0.48	0.49	0.49	0.49
Residual Std. Error	0.54	1.01	1.91	0.54	1.00	1.89	0.53	0.98	1.88
Out-of-sample R ²	0.38	0.43	0.60	0.39	0.43	0.59	0.41	0.46	0.58
<i>Panel B. 1992-2009</i>									
Staff Forecast	0.81*** (0.21)	0.31 (0.22)	0.06 (0.20)	0.79*** (0.19)	0.25 (0.21)	-0.17 (0.21)	0.77*** (0.18)	0.15 (0.24)	-0.34 (0.27)
Tonality				0.03 (0.06)	0.08 (0.12)	0.40** (0.18)			
Trend Tonality							0.05 (0.09)	0.25 (0.21)	0.73** (0.35)
Tonality Shock							-0.001 (0.05)	-0.12 (0.13)	0.05 (0.12)
Intercept	0.14 (0.17)	0.88*** (0.31)	2.34*** (0.53)	0.14 (0.17)	0.92*** (0.27)	2.70*** (0.43)	0.13 (0.17)	0.95*** (0.26)	2.92*** (0.48)
P(Forecast = 1)	0.37	0.00	0.00	0.26	0.00	0.00	0.20	0.00	0.00
Observations	144	144	144	144	144	144	144	144	144
Adjusted R ²	0.20	0.02	-0.01	0.20	0.03	0.13	0.19	0.06	0.19
Residual Std. Error	0.46	0.77	1.09	0.47	0.77	1.01	0.47	0.76	0.97
Out-of-sample R ²	0.66	0.69	0.80	0.65	0.69	0.78	0.63	0.65	0.77

Notes: Estimates from the regression of 1-,2- and 4- quarter inflation on Fed Staff forecast of inflation, and Tonality (or Trend and Shock components of Tonality). Panel A shows estimates between 1970 to December 1991; Panel B shows estimates after January 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 R² 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 9: Economic Forecast Regressions conditional on Blue Chip Forecasts: 1992-2009

Quarters Ahead	GDP		Unemployment		Inflation	
	2	4	2	4	2	4
Forecast	0.91*** (0.19)	0.69*** (0.25)	1.09*** (0.14)	0.98*** (0.18)	-0.33 (0.44)	-0.54 (0.40)
Trend Tonality	0.45* (0.24)	1.12** (0.47)	-0.27** (0.14)	-0.66** (0.31)	0.35 (0.26)	0.82** (0.40)
Tonality Shock	0.10 (0.13)	0.11 (0.20)	0.05 (0.05)	-0.03 (0.08)	-0.14 (0.14)	0.04 (0.13)
Intercept	0.17 (0.46)	0.67 (1.02)	0.07 (0.12)	0.32 (0.29)	1.50*** (0.41)	3.49*** (0.87)
P(Forecast = 1)	0.66	0.23	0.51	0.91	0.00	0.00
Observations	144	144	144	144	144	144
Adjusted R ²	0.50	0.38	0.72	0.61	0.07	0.20
Residual Std. Error	0.99	1.62	0.43	0.76	0.75	0.97
Out-of-sample R ²	0.50	0.38	0.69	0.57	0.31	0.47

Notes: Estimates from the regression of 2- and 4- quarter cumulative GDP growth, unemployment rate change and inflation on Blue Chip consensus forecast for the corresponding variable and Trend and Shock components of Tonality for January 1992 to 2009. The first two columns show GDP growth rate regression estimates, the next two show change in unemployment rate regression estimates, and the last two column show inflation 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 R² 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: Regressions Predicting Fed Funds Rate Change Change: 1992-2009

Quarters Ahead	1	2	4	1	2	4	1	2	4
Staff Forecast	1.19*** (0.05)	1.34*** (0.11)	1.51*** (0.20)	1.15*** (0.05)	1.27*** (0.12)	1.39*** (0.22)	1.15*** (0.05)	1.27*** (0.12)	1.37*** (0.22)
Tonality				0.07** (0.03)	0.12* (0.07)	0.27* (0.14)			
Trend Tonality							0.07 (0.05)	0.14 (0.11)	0.32 (0.22)
Tonality Shock							0.06 (0.04)	0.09 (0.09)	0.22 (0.20)
Intercept	-0.09** (0.04)	-0.21** (0.10)	-0.43 (0.27)	-0.13*** (0.04)	-0.27** (0.11)	-0.55* (0.29)	-0.13*** (0.05)	-0.28** (0.12)	-0.57** (0.28)
P(Forecast = 1)	0.00	0.00	0.01	0.00	0.02	0.07	0.00	0.03	0.09
Observations	144	144	144	144	144	144	144	144	144
Adjusted R ²	0.85	0.70	0.47	0.85	0.71	0.49	0.85	0.71	0.48
Residual Std. Error	0.34	0.68	1.34	0.34	0.67	1.31	0.34	0.67	1.32
Out-of-sample R ²	0.85	0.69	0.47	0.85	0.70	0.48	0.85	0.69	0.46

Notes: Estimates from regressions of Fed funds rate change over the next 1-, 2- and 4-quarters using Fed staff forecasted change in Fed funds rate, Tonality, and its Trend and Shock components. 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 R² 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 11: Regressions Predicting Fed Funds Rate Change Change with Blue Chip Forecasts: 1992-2009

Quarters Ahead	1	2	4	1	2	4	1	2	4
Blue Chip Forecast	1.17*** (0.04)	1.35*** (0.08)	1.55*** (0.20)	1.12*** (0.04)	1.29*** (0.08)	1.44*** (0.19)	1.12*** (0.04)	1.28*** (0.09)	1.42*** (0.18)
Tonality				0.08*** (0.03)	0.11** (0.05)	0.31** (0.12)			
Trend Tonality							0.08* (0.04)	0.14 (0.09)	0.52** (0.21)
Tonality Shock							0.08** (0.04)	0.07 (0.08)	0.04 (0.16)
Intercept	-0.10*** (0.03)	-0.25*** (0.09)	-0.71*** (0.27)	-0.14*** (0.04)	-0.30*** (0.09)	-0.82*** (0.24)	-0.14*** (0.04)	-0.31*** (0.09)	-0.92*** (0.23)
P(Forecast = 1)	0.00	0.00	0.01	0.00	0.00	0.02	0.00	0.00	0.02
Observations	144	144	144	144	144	144	144	144	144
Adjusted R ²	0.88	0.76	0.56	0.89	0.77	0.59	0.89	0.77	0.60
Residual Std. Error	0.30	0.60	1.22	0.29	0.59	1.18	0.29	0.59	1.16
Out-of-sample R ²	0.89	0.77	0.56	0.89	0.78	0.58	0.89	0.78	0.59

Notes: Estimates from regressions of Fed funds rate change over the next 1-, 2- and 4-quarters using Blue Chip consensus Fed fund forecasted change, Tonality, and its Trend and Shock components. 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 R² 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 Fed Funds Rate Change with Term Spread: 1992-2009

Quarters Ahead	Staff Forecast			Blue Chip Forecast		
	1	2	4	1	2	4
Forecast	0.82*** (0.06)	0.84*** (0.16)	0.59 (0.37)	0.85*** (0.04)	1.00*** (0.11)	0.99*** (0.21)
Tonality	0.04* (0.02)	0.08 (0.06)	0.17 (0.13)	0.05** (0.02)	0.08 (0.06)	0.22* (0.12)
Revision	0.02 (0.07)	0.07 (0.26)	0.20 (0.52)	0.09 (0.08)	0.07 (0.20)	0.10 (0.43)
Term Spread	0.86*** (0.17)	0.82*** (0.26)	1.59*** (0.61)	0.69*** (0.12)	0.54** (0.25)	1.00** (0.48)
Intercept	-0.16*** (0.04)	-0.35*** (0.11)	-0.62** (0.27)	-0.16*** (0.03)	-0.34*** (0.11)	-0.79*** (0.26)
P(Forecast = 1)	0.00	0.33	0.27	0.00	0.98	0.94
Observations	144	144	144	144	144	144
Adjusted R ²	0.88	0.75	0.56	0.91	0.78	0.61
Residual Std. Error	0.30	0.62	1.22	0.26	0.57	1.14
Out-of-sample R ²	0.88	0.74	0.52	0.92	0.80	0.61

Notes: Estimates from regressions of Fed funds rate change over the next 1-,2- and 4-quarters using Fed fund forecasted change, Tonality, revision to the Fed funds forecast and Term spread. Term spread is the difference between Fed fund futures rate and Fed funds target rate at the time of the Greenbook forecast. The first three columns show the estimates with Fed staffs forecasted fed funds rate change; the last three columns show the estimates with Blue Chip consensus forecasted Fed funds rate change. 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 R² 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 13: 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	1.81** (0.78)	3.63** (1.65)	5.59* (2.96)	2.28*** (0.75)	4.45*** (1.61)	7.19** (3.03)			
Tonality Shock	-0.55 (0.63)	-0.50 (1.06)	1.26 (1.48)						
Current Unemp				1.10*** (0.35)	1.81*** (0.69)	2.94* (1.58)	0.83** (0.34)	1.29* (0.66)	2.09 (1.51)
Intercept	0.03 (0.53)	0.10 (1.06)	-0.08 (2.15)	-6.83*** (2.32)	-11.19** (4.57)	-18.40* (10.62)	-5.12** (2.24)	-7.85* (4.43)	-13.00 (10.60)
Observations	358	358	358	358	358	358	358	358	358
Adjusted R ²	0.021	0.041	0.054	0.056	0.086	0.111	0.020	0.022	0.029
Residual Std. Error	7.803	11.541	16.438	7.660	11.266	15.942	7.806	11.653	16.660
Out-of-sample R ²	0.007	0.012	0.046	-0.021	0.028	0.047	-0.049	-0.010	-0.025

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 eighth prospective 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 R² shows fit of S&P 500 returns from the prediction regression versus the historical mean. The out-of-sample R² are calculated over the period June 1975 through December 2009. *p<0.1; **p<0.05; ***p<0.01

Table 14: Determinants of Recession in a Probit Regression: 1973-2009

Months Ahead	3	3	6	6	12	12
Trend Tonality	-0.94*** (0.33)	-0.53* (0.32)	-1.03*** (0.38)	-0.83** (0.34)	-0.72* (0.37)	-0.82*** (0.25)
Tonality Shock	-0.23 (0.20)	-0.21 (0.19)	-0.08 (0.17)	0.02 (0.14)	-0.18 (0.11)	-0.14 (0.11)
Rec. 2 Months Ago	0.97** (0.41)	1.65*** (0.52)	-0.48 (0.41)	-0.14 (0.51)	-1.27** (0.56)	-0.74 (0.57)
RGDP Forecast	-0.57** (0.25)	-0.52* (0.27)	-0.16 (0.20)	0.002 (0.19)	-0.19 (0.20)	-0.27* (0.16)
Unemp Forecast	-0.87 (0.59)	-1.01 (0.64)	0.21 (0.57)	0.21 (0.50)	0.05 (0.56)	-0.92*** (0.34)
Term Spread		-0.80*** (0.22)		-0.64*** (0.17)		-0.82*** (0.29)
GZ Spread		0.69*** (0.26)		0.46** (0.20)		0.49* (0.26)
Intercept	-0.70* (0.39)	-1.61*** (0.62)	-0.80* (0.47)	-1.56** (0.63)	-0.28 (0.72)	-0.22 (0.69)
Pseudo R2	0.42	0.57	0.27	0.40	0.20	0.36
Observations	323	323	323	323	312	312

Notes: Estimates from Trend and Shock components of Tonality in Probit models of Recession over next 3-, 6- and 12-months. For each horizon we control for elements of staff forecast for unemployment change and GDP growth rate for the matching horizon (1-, 2-, or 4-quarters ahead). We also control for Term spread (10-year minus 1-year Treasury yield) and the GZ spread (Gilchrist and Zakrajšek 2012). *p<0.1; **p<0.05; ***p<0.01

Table 15: Quantile Regressions Predicting 4-Quarter Errors: 1972-2009

	Unemployment		GDP			Fed Funds	
	Tonality	Trend	Tonality	Tonality	Trend	Tonality	Tonality
Q_{90}	-0.47 (0.07)	-0.82 (0.10)	0.21 (0.16)	-0.17 (0.19)	0.39 (0.15)	0.65 (0.18)	
Q_{75}	-0.28 (0.06)	-0.48 (0.09)	0.28 (0.17)	0.38 (0.22)	0.34 (0.18)	0.53 (0.22)	
Q_{50}	-0.13 (0.05)	-0.23 (0.07)	0.33 (0.17)	0.87 (0.26)	0.23 (0.17)	0.20 (0.21)	
Q_{25}	-0.03 (0.05)	-0.12 (0.07)	0.51 (0.16)	1.15 (0.25)	0.60 (0.16)	0.78 (0.25)	
Q_{10}	0.00 (0.05)	0.05 (0.07)	0.75 (0.15)	1.60 (0.26)	0.57 (0.15)	0.69 (0.25)	
$P(Q_{50} = Q_{90})$	0.005	0.000	0.333	0.085	0.739	0.399	
$P(Q_{50} = Q_{10})$	0.263	0.063	0.302	0.178	0.305	0.606	

Notes: Estimates from the quantile regressions of 4- quarter cumulative GDP growth, change in unemployment rate and change in Fed Funds rate errors on Tonality (or Trend and Shock components of Tonality). Table shows the estimates for July 1972 to December 2009 in columns 1 through 4 and January 1992 to December 2009 for columns 5 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 *Forthcoming*).

Table 16: Quantile Regressions Predicting Excess S&P 500 Returns over 3,6,12 Months: 1970-2009

Months Ahead	Trend Tonality		
	3	6	12
Q_{90}	-0.81 (0.71)	-1.46 (1.04)	0.12 (1.14)
Q_{75}	0.11 (0.73)	0.50 (1.16)	2.61 (1.45)
Q_{50}	1.47 (0.80)	3.38 (1.20)	4.06 (1.83)
Q_{25}	3.44 (0.72)	6.70 (1.15)	9.80 (2.10)
Q_{10}	4.38 (0.71)	8.47 (1.15)	13.86 (2.03)
$P(Q_{50} = Q_{90})$	0.025	0.01	0.132
$P(Q_{50} = Q_{10})$	0.053	0.01	0.010

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), and the fourth prospective meeting (six months hence). Table shows the estimates after January 1970 to December 2009. 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 *Forthcoming*).

Table 17: Predictive content of Positive and Negative Tonality: 1990-2009

	Quarters ahead				Months ahead	
	GDP		Unemployment		Excess Returns	
	2	4	2	4	3	6
Forecast	0.78*** (0.14)	0.59*** (0.17)	1.05*** (0.14)	1.00*** (0.18)		
Trend Positivity	0.79** (0.31)	1.89*** (0.59)	-0.21 (0.14)	-0.70** (0.35)	3.36** (1.64)	5.90* (3.17)
Trend Negativity	0.15 (0.35)	0.29 (0.55)	0.40** (0.19)	0.58* (0.34)	0.41 (2.31)	-0.29 (5.18)
Intercept	-0.06 (0.50)	-0.14 (0.91)	-0.02 (0.15)	0.38 (0.38)	-2.49 (1.84)	-3.94 (3.66)
P(Forecast = 1)	0.04	0.01	0.29	0.72	0.20	0.35
Observations	144	144	144	144	144	144
Adjusted R ²	0.52	0.48	0.72	0.61	0.05	0.07
Residual Std. Error	0.97	1.50	0.42	0.76	7.35	11.49
Out-of-sample R ²	0.48	0.34	0.71	0.56	0.01	0.03

Notes: Estimates of predictive content of the Trend and Shock components of Positive and Negative Tone when controlled for staff forecast for GDP growth rate, unemployment rate changes. Standard errors shown below coefficient estimates are corrected for autocorrelation using the automatic bandwidth selection procedure described in (Newey and West 1994). For GDP and unemployment rate changes ($2 \cdot k + 1$) lags are use where k is the forecast horizon; for three and six months ahead excess S&P 500 returns, the respective lags are 1 and 3. The out-of-sample R² are calculated over the period that begins 64 meetings into the start of the sample through December 2009. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$