

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

The Power of Narratives in Economic Forecasts

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

2020-001

Please cite this paper as:

Sharpe, Steven A., Nitish R. Sinha, and Christopher A. Hollrah (2020). “The Power of Narratives in Economic Forecasts,” Finance and Economics Discussion Series 2020-001r1. Washington: Board of Governors of the Federal Reserve System, <https://doi.org/10.17016/FEDS.2020.001r1>.

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.

The Power of Narratives in Economic Forecasts

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

First draft: August 30, 2017
Current draft: November 5, 2020

Abstract

The sentiment, or “Tonality”, extracted from the narratives that accompany Federal Reserve economic forecasts is strongly correlated with future economic performance, positively with GDP and negatively with unemployment and inflation. Moreover, Tonality conveys incremental information in that it predicts errors in both Federal Reserve and private-sector forecasts of GDP, unemployment, and monetary policy up to four quarters out. Tonality similarly predicts stock returns. Tonality is most informative when uncertainty is high and point forecasts predict subpar growth. Quantile regressions indicate that much of Tonality’s forecasting power arises from its signal of downside risks to economic performance and stock returns.

JEL codes: E17, E52, G14.

Keywords: Text Analysis, Economic Forecasts, Monetary Policy, Stock Returns

* Sharpe (Steve.A.Sharpe@frb.gov) and Sinha (Nitish.R.Sinha@frb.gov) are in the Research and Statistics division at the Federal Reserve Board, 20th Street and Constitution Avenue, NW, Washington DC 20551; Hollrah (Chollrah@umich.edu) is at the University of Michigan. Our views do not necessarily reflect those of the Federal Reserve System or its Board of Governors. We are very grateful for the research assistance provided by Toby Hollis, Taryn Ohashi, and Stephen Paolillo. Many thanks to Jeremy Rudd for his help in developing the wordlists. We are also thankful to our Board colleagues Jack Bao and Neil Ericsson for detailed discussions. An early version of this paper was circulated as “What’s the story: A New Perspective on the Value of Economic Forecasts.”

I. Introduction

Over the years, even as many researchers and market participants have questioned the value of macroeconomic forecasts, substantial resources continue to be devoted to their production and dissemination. For instance, the Blue Chip Survey of Economic Indicators collects monthly 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 of success in predicting financial prices. Similarly, eight times a year, prior to each meeting of the FOMC committee, the staff at the Federal Reserve Board provide a detailed forecast of the U.S. economy (staff forecast). Our study provides a new perspective on the information embedded in macroeconomic forecasts and their potential value to policymakers and financial market participants.

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

This study breaks new ground by applying tools from the emerging literature on textual analysis to try and gauge the incremental value of any signal conveyed by the sentiment extracted from the narratives that accompany forecasts. To do so, we focus on Federal Reserve Board forecasts, which are described in the Greenbook and are perhaps the longest available time series of macroeconomic forecasts for the U.S. economy. In particular, we quantify the degree of optimism versus pessimism embedded in the forecast narrative, which we call the “Tonality”

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.

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 measure of forecast narrative sentiment that we extract is found to be strongly correlated with direction of the accompanying point forecasts for key economic variables, usually with the intuitive sign. In particular, Tonality is positively correlated with forecasts of GDP growth and negatively correlated with the forecast trajectory of unemployment and inflation.

The central question we consider is whether, to what extent, and why, such a measure of text sentiment has value as a signal of future economic performance. To answer this, we examine whether Tonality has incremental power, over and above the point forecasts, for predicting key macroeconomic quantities—namely unemployment, GDP growth, and inflation. We pursue the hypothesis that positive sentiment helps to predict more favorable economic outcomes, such as higher GDP growth. In particular, we estimate both OLS and quantile regressions of Greenbook forecast errors on Tonality.

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

What is more, the quantile regressions reveal strong asymmetry in Tonality’s predictive content. For both GDP growth and the unemployment rate, and at all horizons, the results indicate that Tonality is most informative about the likelihood of bad economic news; that is, it provides a particularly strong signal of lower tail risks for GDP growth, relative to forecast, and signals the prominence of upper tail risks for unemployment relative to forecast. In contrast, we find that while Tonality does not contain a directional signal for inflation, lower Tonality signals larger tail risks to inflation forecasts in either direction.

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

One possible explanation for Tonality’s predictive power could be stickiness in the Greenbook point forecasts, that is, forecast revisions that are more sluggish than would be optimal for minimizing mean square error. Nordhaus (1987) first described this as “Inefficient forecasts ... let the news seep in slowly” and argues that the resultant forecasts errors would be predictable, in part, using recent forecast revisions.² Such an inefficiency in Greenbook point forecasts could account for the predictive power of the narrative if the sentiment in the narrative is simply more “nimble” to incorporate new information. We find little evidence that sticky forecasts can account for much of Tonality’s predictive power.

To get a better sense of the nature of the information conveyed by Tonality, and when it might be most useful, we examine whether its predictive power is stronger when macroeconomic uncertainty is high, or, similarly, when the staff GDP forecast calls for below-trend growth. One long-perceived weakness of economic forecasts, documented early on by Zarnowitz and Braun (1993) and revisited recently by Smirnov and Avdeeva (2016), is that point forecasts rarely call for an outright decline in GDP before a recession has actually begun. Thus, a plausible hypothesis is that the information value of Tonality is higher when forecasts call for subpar growth; at such times, the forecast narrative might convey reasons for, or the risks surrounding, a

² More recently, in an analysis of consensus forecasts from the Survey of Professional Forecasters, Coibion, and Gorodnichenko (2015) find evidence of “information rigidity,” in that forecast revisions for inflation tend to predict future forecast errors in the same direction. Dovern, et al. (2015) go even further, showing that individual forecast revisions also tend to predict an individual forecasters errors in the same direction, though the magnitude of rigidity is smaller than in consensus forecasts.

subpar outlook. Indeed, we show that, when uncertainty is high or when the GDP forecast calls for sub-par growth, the predictive power of point forecasts is remarkably poor; at the same time, Tonality conveys a relatively strong signal about the balance of risks to the forecast.

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

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

Finally, we examine whether Tonality predicts stock market returns. We test whether the information embedded in Tonality would have conveyed valuable information for investors. In particular, higher tonality predicts a greater likelihood of stronger future economic outcomes; if that signal has not already been incorporated by investors, then one might expect higher Tonality to predict higher stock returns. On the other hand, given our finding that higher Tonality also tends to signal news of tighter monetary policy, the tighter monetary policy predicted by higher Tonality could temper any positive stock market effect related to the favorable macroeconomic information that Tonality conveys.

Nonetheless, we find that Tonality has substantial power for predicting excess returns on stocks over holding periods ranging from 3 to 12 months following the production of the Greenbook. The positive coefficient we find on Tonality is consistent with the hypothesis that its predictive power for stock returns arises from its ability to predict GDP and cash flow news. That is, higher Tonality predicts subsequent news of a stronger economy (and thus cash flows) and presumably lowers investor risk premiums, both of which would boost stock values. Indeed, when we control for the state of the economy as gauged by the unemployment rate at that time of forecast, which is correlated with Tonality and presumably with the equity risk premium, Tonality's predictive power is even stronger. Echoing the findings for predicting economic performance, the predictive power of Tonality is highest when the growth forecast is subpar. What is more, Tonality appears to magnify downside risk in excess returns.

Of course, Tonality, unlike the standard conditioning variables used in that literature, is not directly observable by investors; however, this does beg the question of whether and when Tonality might be conveyed indirectly to the public. Thus, before concluding, we ask whether the sentiment gauged by Greenbook Tonality might be transmitted to the public in one of the two subsequent formal FOMC communications, the FOMC statement released following the FOMC meeting and the FOMC meeting minutes released several weeks hence. We find that the Tonality of the relatively terse FOMC statements appear to convey little of that sentiment; in contrast, Tonality measured from the FOMC minutes correlates noticeably with Tonality in the corresponding (preceding) Greenbook. This suggests that a more careful analysis of the sentiment conveyed in the FOMC Minutes Tonality might find this transcript to have some of the forecasting properties of Greenbook Tonality.

While adding to the literature on the efficacy of economic forecasts, our study also contributes to the relatively new and burgeoning line of research in economics that draws insights from treating text as a new source of data. Perhaps most closely related is the nascent research in economics and finance that attempts to quantify narratives, an agenda just recently nudged into the mainstream by Shiller's (2017) presidential address to the American Economic Association. In particular, our approach somewhat echoes studies that examine whether the tone of newspaper articles helps explain or predict stock market returns, beginning with Tetlock (2007), using techniques elaborated upon more recently, for instance, by Garcia (2013), Heston

and Sinha (2017) , Calomiris and Mamaysky (2019) and Ke, Kelly and Xiu (2019). It also bears similarities to Asquith, Mikhail and Au (2005), which examines how the sentiment of the text in Wall Street analyst reports explains firms' stock price responses to earnings forecast revisions. Perhaps it is closest in spirit is Jones, Sinclair and Stekler (2019), which manually score the narrative contained in Bank of England inflation reports, and uses that metric to help predict quarter-ahead inflation.

Also related are recent studies that quantify information conveyed in monetary policy communications and characterize its impacts on markets. Hansen and McMahon (2016) attempt to parse FOMC statements into the information conveyed about either forward guidance or economic conditions and find that the forward guidance has more noticeable market impact. Hansen and McMahon (2017) use text analysis to infer change in the nature of FOMC deliberation following increased transparency. Schmeling and Wagner (2017) gauge the tone of European Central Bank press conferences and find that a more positive tone induces higher interest rates and lower credit spreads and equity volatility. Carvalho, Hsu and Nechio (2016) use sentiment quantified from FOMC communications to examine interest rate reactions to FOMC communication during the zero lower bound period Our study differs from these in that we focus on sentiment embedded in the communications between Fed staff and the FOMC committee, information that is only available to the public years later.

The paper also speaks to the measurement of time-varying macroeconomic uncertainty. Baker, Bloom and Davis (2016) count uncertainty related words in newspapers. Jurado, Ludvigson and Ng (2015) measure the predictability of macroeconomic variables from all available data including financial variables to construct a measure of uncertainty. Relatedly, Clark, McCracken and Mertens (2020) construct an uncertainty measure using the forecast errors of macroeconomic forecasters. We find that forecast narratives have information over and above these uncertainty measures. Moreover, we also observe that the forecast narrative is most informative when macroeconomic uncertainty is high or, similarly, when the forecasted level of growth are unusually low.

Section II describes how we measure Tonality and explores how it co-varies with the point forecasts of key macroeconomic variables in the Greenbook. In section III, we examine the extent to which Tonality conveys information about future macroeconomic conditions not

already reflected in point forecasts. Section IV examines the relevance of the information in Tonality for market participants, beginning with its ability to predict errors in the Blue Chip consensus forecasts. It then examines Tonality’s ability to signal for future monetary policy surprises and stock returns. Finally, it briefly examines whether Greenbook Tonality is transmitted to the public in either the post-meeting FOMC statements or the FOMC meeting minutes. Section V concludes.

II. Measurement of Tonality in Greenbook Text

A. Measuring Tonality

Prior to every scheduled FOMC meeting, Federal Reserve Board staff puts together its forecast for the U.S. economy in an internal Fed document called the *Greenbook* (now the *Tealbook*), which is made public after a 5-year lag. Greenbook forecasts were produced monthly until 1981; thereafter, the frequency dropped to eight per year. Our sample begins January 1970, shortly after the staff’s quantitative quarterly forecast began to look forward more than two quarters. For most of our sample, text analysis is based on the text of Greenbook Part 1, the Summary and Outlook, which outlined the forecast. Prior to the document’s restructuring in August 1974, we analyze text from the section titled Recent Developments and Outlook for Domestic Economic Activity. Our sample ends in December 2009, the last full year before Greenbook was replaced by Tealbook A, which consolidated Greenbook with some closely related content from the also-retired Bluebook.

We construct an index that quantifies the optimism and pessimism of the Greenbook text, which we refer to as “Tonality.” Tonality is equal to the difference between the weighted sum of positive and negative words from our word list. To classify words as “positive” or “negative,” we create a custom dictionary of 231 positive words and 102 negative words.³ To derive our dictionary, we adopt the initial classification of positive and negative words in the widely used Harvard psycho-social dictionary⁴ but then exclude words that have a different connotation in the forecasting context. For example, in contrast to the psycho-social dictionary, we do not consider the words “demean” or “hedge” as negative. Positive words in our dictionary include terms like

³ For the list of positive and negative words, see the online publication appendix A.

⁴ Tetlock (2007) used Harvard-Psychosocial dictionary to quantify the sentiment in financial news. Da, Engelberg and Gao (2014) use Google searches on select words from this dictionary to quantify fear among U.S. investors.

“enthusiasm,” “abundant,” “enhance,” and “successful,” whereas examples of negative words include “unrest,” “fragile,” “trouble,” and “gloomy.” Our approach is most similar to Tetlock (2007) and Loughran and McDonald (2011), who examine word frequency without trying to gauge the context in which words are used. Like Tetlock (2007), we use the Harvard IV Psychosocial dictionary to classify words; and, like Loughran and McDonald (2011), we use weighted word counts and we cull from the list any words that have domain-specific connotation in economic forecasts.⁵

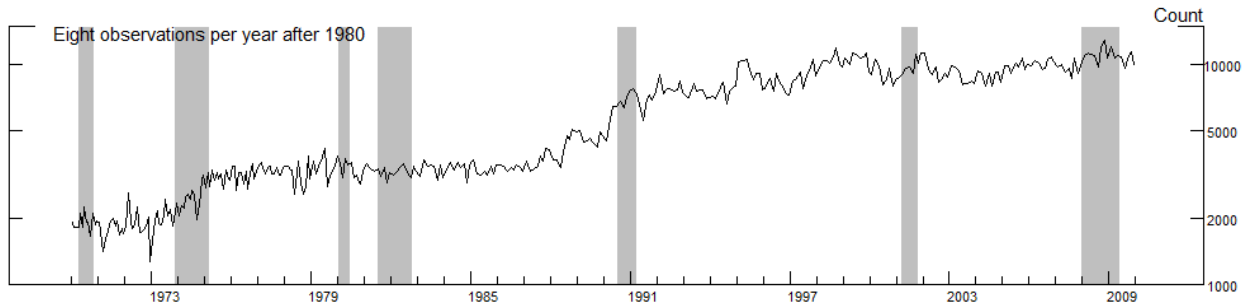
By using the whole document to quantify the overall degree of optimism, irrespective of how words are grouped, we have chosen not to use more elaborate methods of text analysis that would, for instance, attempt to connect the words that convey sentiment with their antecedents, such as particular economic indicators, or which attempts to identify negations.⁶ Such approaches would require a good deal of additional judgment, for instance, on how to classify “nearby” words in text space. It would also necessitate excluding a lot of information such as the descriptors of the many other economic variables that are related to the specific indicators on which we focus.

Figure 1 shows the time series of the total word counts from Greenbook Part I (or its pre-August 1974 equivalent) for our entire sample period. As shown, in the earlier forecast documents, the word count from the outlook section ran at only about 2000 words. After the restructuring in August 1974, the count quickly moved up to about 3000 words, where it hovered until 1990, after which the document gradually ramped up to about 9000 words.

Figure 1: Total words in the Greenbook

⁵ Using the Loughran-McDonald wordlist instead would yield a very different measure of Tonality, which has only a 24 percent correlation with our measure of Tonality in the Greenbook text, although, separately, positive and negative components of the two measures have 78 percent and 81 percent correlations, respectively.

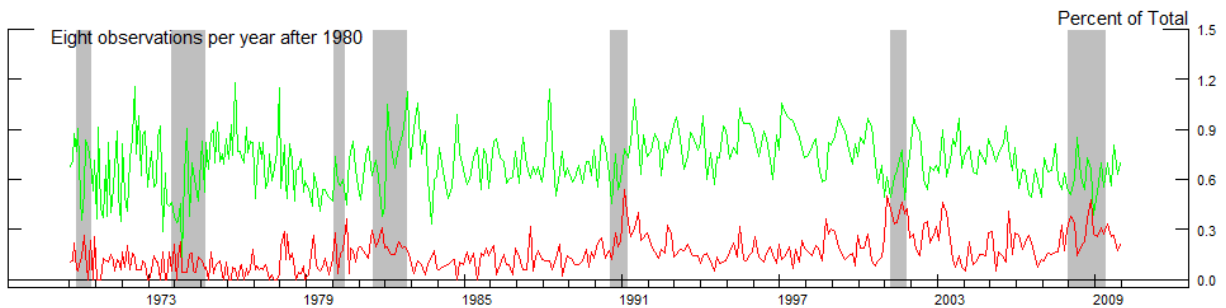
⁶ As one robustness check, we examined sensitivity of our scores to presence of signed words that follow negations. For example, in the clause “GNP is likely to show no further rise”, “rise” follows “no” and should not be counted as a positive word. To examine this, we mute all words in a clause that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



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

Figure 2 shows the number of positive and negative words as a percent of the total word count in each Greenbook. In most documents, the frequency of positive words is far above that for negative words. Also apparent from this picture, prior to the August 1974 restructuring, the percentage of positive words per document appears to have been considerably more variable from one document to the next.

Figure 2: Proportion of Positive and Negative Words in the Greenbook



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

The Tonality index of a document compares the number of positive and negative words in its text, using a weighting scheme in which a word's frequency of appearance in any given Greenbook is normalized by its average frequency in a comparable set of Greenbooks, a

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

The tf-idf weighing scheme is based on the intuition that infrequently used words are especially informative and so receive relatively high weight in the index, whereas very frequently used words are discounted. Common application of tf-idf scheme would have used the inverse document frequency over *all* the Greenbooks. We chose a moving window of roughly five years to account for changes over time in Greenbook writing style. Nevertheless, the correlation between 40-greenbook rolling window tf-idf scores and a simple tf-idf scheme that “sees” all greenbooks is over 95 percent, suggesting the choice of window does not have a substantial effect on our measure of Tonality. Finally, the Tonality index is standardized to have zero mean and standard deviation equal to one. We adapt the Python machine learning library Scikit Pedregosa, et al. (2012) for tf-idf scoring of Greenbooks. Word clouds showing the 50 most prominent positive and negative words in Greenbook during a couple different time periods are shown in the online publication appendix B. Negative words have higher propensity to appear during periods that contain recession.

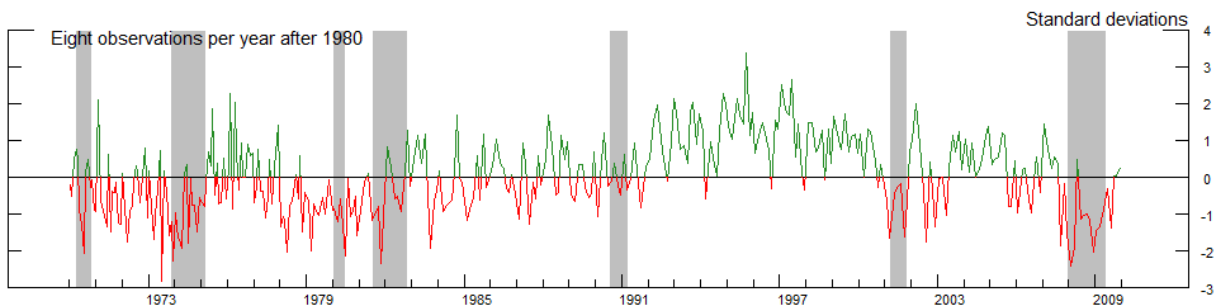
Figure 3 shows the Tonality index plotted over the full sample period, with positive levels indicated in green and negative levels indicated in red. As one might expect, Tonality appears to be procyclical, with the large majority of observations during recessions in negative territory, and a mixture of positive and negative observations during expansionary periods. Among the most deeply negative readings of Tonality are observations in the year leading up to and during the Great recession and the 1974-75 recession. The most noticeable run of highly positive readings was during the mid-1990s. Despite these cyclical tendencies, Tonality also

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

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

appears to be quite volatile, exhibiting much high-frequency movement that is often quickly reversed. To some extent, these fluctuations might reflect noise in our proxy for sentiment.

Figure 3: Greenbook Tonality plotted over time

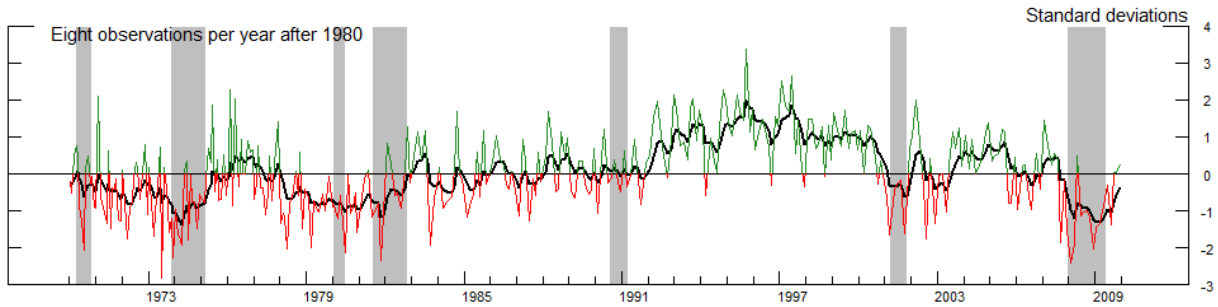


Note: Shaded regions represent NBER-dated recessions. Tonality is standardized to have a zero mean and a standard deviation equal to one. Tonality is shown in green when it is positive and in red when it is negative. Prior to 1981, Greenbooks were produced nearly every month, thereafter the frequency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

Considering that high-frequency movements could reflect noise, or temporary shifts, we construct a smoothed measure of Tonality which is an exponentially weighted moving-average of Tonality. For the post-1980 sample we use a weighting parameter—the decay rate on lagged observations—equal to 0.75; that is, the most recent observation gets a quarter of the weight.⁹ For the pre-1981 sample, when Greenbooks were published at a higher frequency (monthly rather than eight per year), we use a somewhat faster decay rate (0.825), calibrated to imply the same calendar-time decay rate. By construction, smoothed Tonality, which we will call S-Tonality, is meant to reflect the underlying level of Tonality, while deviations from S-Tonality reflect possibly temporary shocks, or “Tonality Shocks.” **Figure 4** shows the resulting time series plot for S-Tonality, along with (raw) Tonality. The cyclical pattern in this smoothed measure of sentiment stands out a bit more clearly. Consistent with our interpretation, the autocorrelation coefficient for the Tonality Shock series is only 0.04.

Figure 4: Greenbook Tonality and trend plotted over time

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

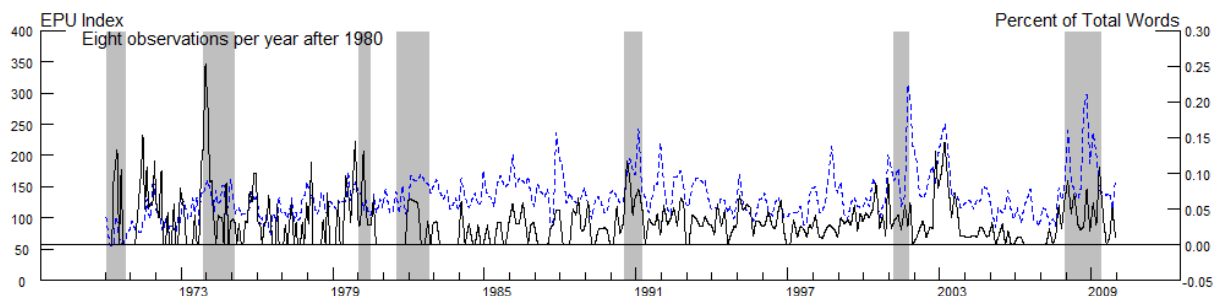


Note: Shaded regions represent NBER-dated recessions. Tonality is standardized to have a zero mean and a standard deviation equal to one. Prior to 1981, Greenbooks were produced nearly every month, thereafter the frequency was reduced to eight times a year. Tonality is shown in green when positive and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

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

An alternative and oft-used metric in text analysis is the extent of uncertainty expressed. In their widely cited study Baker, Bloom and Davis (2016) argue that the frequency of “uncertainty” mentions alongside some key words provides a plausible measure of the degree of uncertainty prevailing with respect to economy, monetary policy, or government policy, called the EPU index. Because the Greenbook, particularly the section we analyze, consists entirely of economic commentary, our adaptation simply involves counting mentions of “uncertainty” and “uncertain” as a fraction of total word count in Part I of each Greenbook. The resulting measure is plotted in **Figure 5**, alongside the time series for EPU. Notably, early in the sample, there are hardly any mentions of uncertainty; and there are relatively few mentions of uncertainty in the run-up to the 2008 financial crisis.¹⁰

Figure 5: Greenbook Uncertainty plotted over time



Note: Shaded regions represent NBER-dated recessions. Prior to 1981, Greenbooks were produced nearly every month, thereafter the frequency was reduced to eight times a year. Instances of ‘Uncertain’ and ‘Uncertainty’ are

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

used to create count of uncertain words, shown as percent of total words (black line), the blue line shows the Baker-Bloom-Davis Economic Policy Uncertainty (EPU) index.

C. Relation of Tonality to Concurrent Greenbook Point Forecasts

To examine whether and how text sentiment is related to the associated point forecast, we examine simple correlations between Tonality and the point forecasts for three key economic performance variables: inflation, the unemployment rate, and GDP growth. The first two constitute the components of the Fed’s “dual mandate.” The third, GDP growth, is perhaps the most frequently cited summary statistic of economic performance, and the quality of GDP forecasts has been extensively studied in the literature. For each economic variable we construct a gauge of the forecast at the one-quarter and four-quarter horizon: for the latter we measure the forecast of cumulative inflation, cumulative GDP growth, and the change in the unemployment rate, each of these over the subsequent four quarters out (with the current-quarter forecast as the base). We also construct the revisions in those forecasts relative to the previous Greenbook. Finally, to gauge the perceived state, or level, of economy activity when the forecast was written down, we use the current-quarter forecast of the unemployment rate.¹¹

The correlations of Tonality and its two subcomponents with four-quarter forecasts, and with revisions to those forecasts are shown in the top section of Table 1. Clearly, most of the correlations with Tonality and S-Tonality are strong, while their signs accord with intuition. Forecast narrative sentiment, as measured by Tonality, is positively correlated with expected GDP growth and negatively correlated with expected inflation and expected changes in unemployment. What is more, all three economic 4-quarter forecasts variables, as well as current-quarter unemployment, are more strongly correlated with S-Tonality than they are with raw Tonality. In contrast, correlations of Tonality and S-Tonality with *revisions* to the GDP and unemployment forecasts are of similar magnitude. And the revisions are the only forecast variables correlated with the Tonality shock, with the intuitive sign, suggesting that some of the

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

volatility in Tonality that is removed from S-Tonality reflects the direction of revisions from the previous forecast.

The lower section of Table 1 shows correlations of Tonality with variables found to help predict economic outcomes or forecast errors in other recent influential studies. It is likely that such variables reflect information similar to that reflected in, and signaled by, Tonality. As shown, Tonality displays only a mild negatively correlation with EPU-Gbk, the measure of uncertainty that results from applying the Baker-Bloom-Davis methodology for EPU to the Greenbook text. Similarly, Tonality has a mild negative correlation with EPU itself. In contrast, Tonality (and S-Tonality even more so) displays a strong negative correlation with MACROU, an economic uncertainty measure devised by Jurado, et al. (2015), and with NFCI, the financial conditions index published by the Federal Reserve Bank of Chicago. Lastly, we find that stock returns over the inter-Greenbook period are moderately positively correlated with Tonality, and S-Tonality. Moreover, positive correlation between Tonality Shock and stock returns indicates that innovations to Tonality are influenced in part by some of the news recently driving stock price changes.

III. Greenbook Tonality as Contributor to Forecast

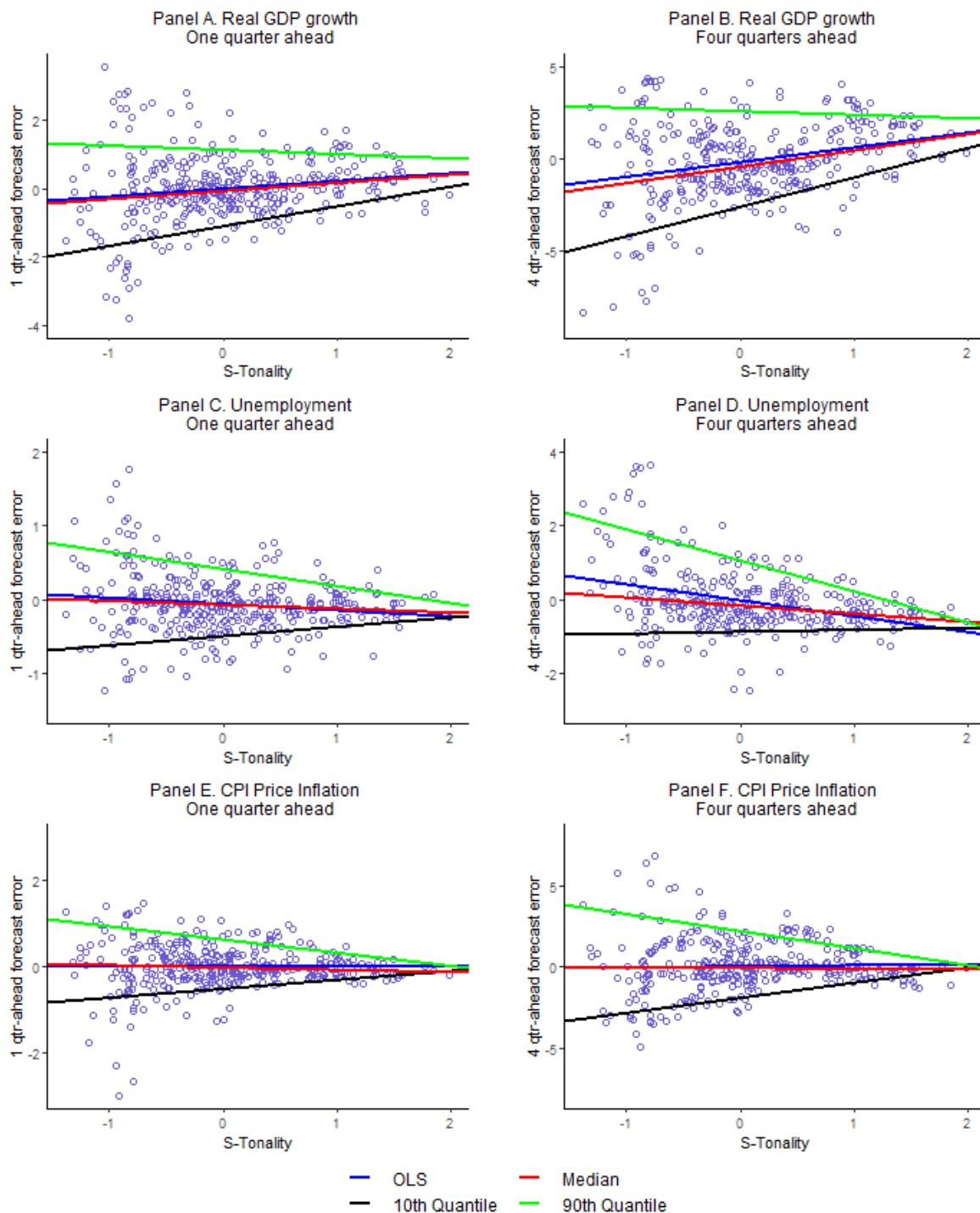
A. Univariate Predictor of Forecast Errors

Having established a strong connection between the point forecasts of key economic performance measures and Tonality of the forecast narrative in the same document, our analysis turns to a central question of interest: does Tonality provide incremental predictive power for such measures of economic performance? In particular, does Tonality contain information regarding future GDP growth or unemployment that is not fully reflected in the GDP or unemployment forecast itself? To gauge the predictive content of Tonality, we estimate regressions that test whether Tonality can predict Fed staff forecast errors. In each regression, the dependent variable is the realized forecast error, while the explanatory variable is either Tonality or S-Tonality of the narrative from the corresponding Greenbook. For GDP, the forecast error is measured relative to 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. More details are provided in the Internet appendix.

For each economic forecast variable and horizon, OLS is used to estimate the conditional expected forecast error as a linear function of Tonality (or S-Tonality). In addition, for each forecast, we use quantile regression to provide estimates of the 10th and 90th quantiles of the forecast error distribution, conditional on Tonality. This allow us to gauge whether Tonality signals information about the tails of the distribution of forecast errors, over and above any effect on the mean forecast error: do downside or upside tail risks to the forecast vary notably with Tonality? Table 2 provides the relevant statistics from each regression, while figure 6 provides a visual representation of the data and the slope coefficients on S-Tonality.

The top two panels of **figure 6** show a scatter plot of forecast errors for GDP growth one quarter ahead and (cumulative) four quarters ahead, plotted against S-Tonality from the associated Greenbook narrative. The blue line in each figure depicts the OLS regression line while the other lines show the estimated 10th, median, and 90th quantiles of the forecast errors conditioned on S-Tonality. While close inspection reveals the OLS lines to be upward sloping for both horizons, the more striking pattern is the asymmetry in the signal conveyed by S-Tonality: the 10th quantile line is more steeply upward sloping than the OLS line, whereas the 90th quantile line appears to be flat. The OLS slope indicates that lower sentiment predicts realizations will tend to fall short of forecast, but the strongest signal from S-Tonality is with regard to the likelihood of large negative forecast errors, gauged by the slope coefficient from the 10th quantile of forecast errors.

Figure 6: Regressions Predicting Forecast Errors: OLS, 10th and 90th Quantiles



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

As shown in table 2, the OLS coefficients on Tonality and S-Tonality are significant though sometimes only marginally so, with R-squared statistics ranging from only 2 to 7 percent. The beta coefficients in the 10th quantile regressions are about twice the size of the respective OLS coefficient, and are significant at the 1 percent level in every case, with R-squared statistics ranging from 7 to 12 percent. The strongest signals appear to be from S-Tonality when regressed on the four-quarter forecast error. Finally, the Tonality coefficients for the 90th quantile are insignificant in every case.

Turning to unemployment forecast errors, the middle 2 panels in figure 6 show analogous results, but with the tail effects reversed, because the upper tail (higher than expected unemployment) here represents bad news regarding economic activity. Also somewhat different, it appears that S-Tonality conveys a much stronger signal for four-quarter forecasts than for one-quarter forecasts. As shown in Table 2, the OLS coefficients on Tonality and S-Tonality are significant at the 1 percent level for the four-quarter forecast, while the 90th quantile coefficients are significant at both horizons. Most strikingly, S-Tonality explains 21 percent of the variation in the 4-quarter forecast errors in the 90th quantile regression.

The results for CPI forecast errors are quite different. First, there is no effect of Tonality or S-Tonality on mean CPI forecast errors, as indicated by the OLS regressions. Second, for forecast errors at both horizons, the quantile regressions indicate that lower Tonality predicts increased downside risk as well as upside risk to inflation, that is, the implied effect of lower Tonality is a higher forecast error variance. Indeed, the last four rows of Table 2 confirms that this pattern is statistically significant. Again, while there is no effect on the mean, lower Tonality or S-Tonality raises both the expected 10th percentile as well as 90th percentile, with the effects both statistically and qualitatively stronger with S-Tonality.

Broadly speaking, these results mirror quite remarkably key findings in Adrian, et al (2019) and particularly Adams, et al. (2020). These studies examine how the conditional distribution of the same three economic variables, varies with financial conditions, as reflected in the Federal Reserve Bank of Chicago's National Financial Conditions Index (NFCI). Using quantile regressions, they find that downside risks to forecasts of GDP and upside risks to

forecasts of unemployment increase substantially with less favorable financial conditions. In addition, they find that the financial conditions index conveys no information for predicting mean inflation forecast errors, but that a worsening of financial conditions boosts both upside and downside risks to inflation forecasts. Similarly, Hennge (2019) studies how the conditional distribution of GDP growth varies with one particularly well-motivated measure of economic uncertainty (MacroUnc) introduced by Jurado, et al. (2015). She finds that, like the NFCI, MacroUnc contains substantial predictive information regarding downside risks to future GDP growth, but little information about upside risks.

B. What Factors Might Tonality Reflect?

The similarity of these empirical relationships to the relationship we find between the forecast narrative sentiment and realized errors to the staff's economic forecasts suggests that Tonality may be strongly influenced by financial conditions and uncertainty, as both these variables were built using financial market indicators and are likely to be correlated. Indeed, this plausibility of this hypothesis is clearly indicated by the correlation reported in Table 1 between each those two variable and Tonality and S-Tonality of about -0.47 and -0.68, respectively. One natural question, then, is whether the conditioning information in Tonality reflects the same information in either MacroUnc or NFCI.

This question, among other hypotheses regarding the nature of Tonality's predictive information, is tackled in forecast error regressions where we simultaneously condition on S-Tonality and other variables, results of which are reported in Table 3. For this analysis we focus on GDP and unemployment forecasts, the variables for which Tonality was found to have directional information. For brevity, we focus on the four-quarter forecasts for those two variables, the horizon for which the explanatory power of Tonality and other conditioning variables is highest. Beginning with four-quarter GDP growth, the first row in each block shows (again) the key univariate regression results, specifically, the coefficient on S-Tonality, its p-value and the regression R-squared from the OLS, 10th, and 90th quantile regressions; subsequent rows show the analogous statistics for the S-Tonality coefficient when we control for a competing candidate predictor.

In the case of GDP forecast errors, S-Tonality by itself is marginally significant with a positive coefficient in the OLS regression and an R-squared of 7 percent. The 10th quantile coefficient is double this magnitude with an R-squared of 12 percent, while it is small and not significant for the 90th quantile. Controlling for MacroUnc, however, substantially alters S-Tonality's marginal effects. Broadly speaking, the asymmetry of its predictive effect disappears, with S-Tonality's coefficient becoming markedly larger in the OLS and the 90th quantile regressions, while shrinking some in the 10th quantile regression, and highly significant in all three cases. This suggests at least some of the asymmetry in the predictive power of Tonality shares a common origin with MacroUnc.

Interestingly, as shown in the subsequent line, controlling instead for NFCI leads to very similar results. Moreover, in either case, regression R-squared statistics are notably higher with the additional regressor. This is consistent with the fact (not shown) that each of these controls are themselves statistically significant in the 10th quantile regressions. In sum, when we allow MacroUnc or NFCI to, in effect, control for downside risk, S-Tonality continues to have marginal predictive power, though now roughly similar across the quantiles. The takeaway is that, despite its sizable correlations with MacroUnc and NFCI, and the commonality of their signals regarding downside risk, the predictive information in Tonality is, at least in part, distinct from that signaled by these two variables.

The other two controls—the inter-Greenbook period stock return and the Greenbook forecast revision, considered separately in the subsequent two lines—are meant to serve as signals of recently received information, which staff might have been slow to incorporate into point forecasts. If the explanatory power of S-Tonality resulted largely due to the forecast being sluggish to adjust but the narrative being more nimble, then including the forecast revision or the recent stock market return might diminish the marginal explanatory power of S-Tonality. In both cases, we find coefficients on S-Tonality in the 10th quantile regression to be only a bit smaller, and they remain highly significant. At the same time, we find that including the stock return boosts the predictive power of these regressions, indicating that the forecast does not entirely embed the information signaled by the inter-Greenbook stock return.

For unemployment forecast errors, we find fairly analogous results. Controlling for MacroUnc in the Unemployment regressions reduces the magnitude of the still-significant

negative coefficient on S-Tonality in the 90th quantile regression, and results in a negative coefficient in the 10th quantile regression, thereby diminishing the asymmetry. On the other hand, controlling for the NFCI has less effect on S-Tonality coefficients. As with GDP forecast, controlling for inter-Greenbook stock return nor the Forecast Revision has much effect on the on S-Tonality's coefficient estimate, though including the stock return again boosts overall regression predictive power.

C. When is Tonality most informative?

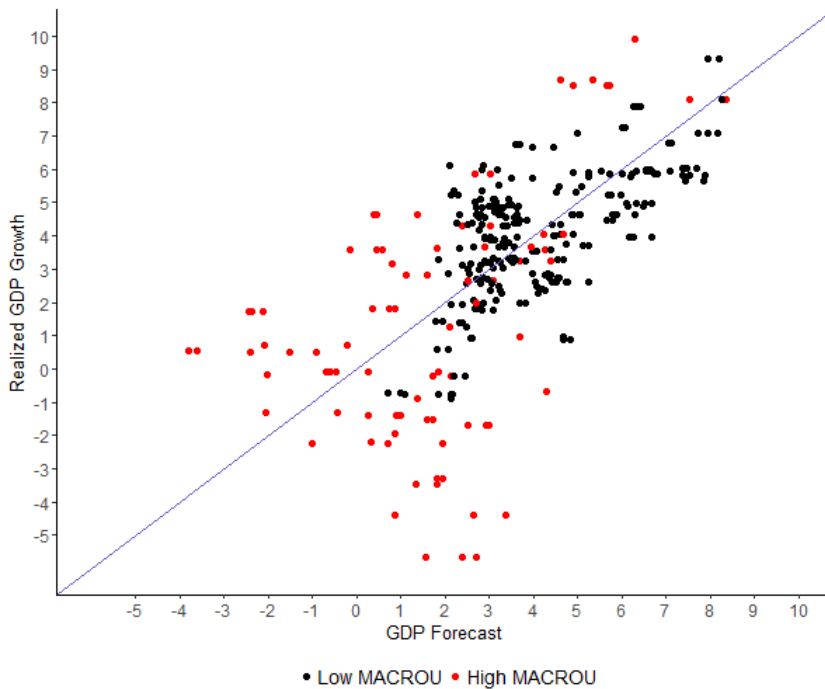
For forecasts of GDP growth and unemployment, the evidence indicates that Tonality is informative both about the mean expected outcome, but perhaps even more so, about downside risks to economic activity. It would be instructive to also determine the ex ante conditions under which Tonality is likely to be most informative. In particular, it would be useful to examine whether Tonality is more informative about the likely direction of forecast errors when, for instance, uncertainty is relatively high. First, some perspective on how GDP forecasts and the associated forecast errors are related to macroeconomic uncertainty, as gauged by MacroUnc, is provided by Figure 7. The scatter plot shows realized four-quarter GDP growth (vertical axis) plotted against the associated forecast published in Greenbook four quarters earlier.

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

We consider the following hypotheses: Does Tonality convey more information when MacroUnc is high (in its top quartile)? Similarly, does tonality convey more information when the GDP forecast calls for subpar growth, compared to other times? While these hypotheses

appear potentially somewhat redundant given our observations from the figure, an inference which conditions on the GDP forecast has the attraction that the forecast is plainly observable (to the FOMC committee), whereas MacroUnc is a construct estimated from data that is not all available in real time. These hypotheses are examined by regressing realized four-quarter economic performance, either GDP growth or the change in unemployment, on the respective point forecast and on S-Tonality. This specification is more general than the forecast error regressions in that it allows a parsing of the Greenbook predictive power between the point forecast and the tone of narrative sentiment.

Figure 7: Realized four-quarter GDP Growth versus Forecast



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

The top section of Table 4 show results for GDP growth. The first two columns show estimates for the low-uncertainty subsample, while the latter two show estimates for the high-uncertainty subsample. For the low-uncertainty sample, the coefficient on the forecast in a

univariate regression is 0.71, significantly below the 1.0 hypothesized in standard rationality tests, while the regression R-squared is 0.40. When S-Tonality is added to the regression, its coefficient is positive and also statistically significant and the R-squared is boosted moderately to 0.47. For the high-uncertainty sample, the coefficient on the forecast in the univariate regression is also 0.71, but here the R-squared is only 0.20. Even more notable, when S-Tonality is added to the regression (in column 4), the forecast coefficient is no longer significant; at the same time, the coefficient on S-Tonality is large and highly significant, while the regression R-squared jumps to 0.37, only moderately below that in the low-uncertainty sample. These results clearly indicate that, when uncertainty is high, the narrative sentiment becomes the more informative indicator of future economic performance.

The lower section of Table 4 shows the analogous results for the unemployment regression. Here, consistent with rationality, the coefficient on the forecasted change in unemployment in univariate regressions is close to unity, in both the low- and high-uncertainty subsamples. Again, however, the predictive content of the forecast as measured by R-squared is more than twice as high in the low uncertainty sample, 0.60 versus 0.48. Moreover, while adding S-Tonality to the regression in the low-uncertainty sample marginally improves predictive power, doing so in the high-uncertainty sample doubles predictive power; and, as in the comparable high-uncertainty GDP regression, the forecast itself no longer contributes any marginal predictive power.

Table 5 shows the analogous set of regression estimates when the sample is conditioned on whether or not the four-quarter point forecast is calling for subpar growth, or $E(\text{GDP}) < 2.5\%$. The top section shows regressions predicting GDP growth. Here, the results are similar to, if not more striking than, those in Table 4. In the first column, the univariate GDP growth regression, in the sample where the GDP forecast exceeds 2.5%, the coefficient on the GDP forecast is 0.77 with an R-squared of 0.31. In this subsample, adding S-Tonality boosts the R-squared to 0.39, while leaving the coefficient on the forecast near unity and highly significant. In contrast, in the third and fourth regression, when the GDP forecast calls for subpar growth, the coefficient estimate on the forecast is only 0.22 and insignificant, while the R-squared is near zero. Adding S-Tonality boosts the regression R-squared to 0.22, owing to the positive predictive power of Tonality.

Finally, as shown in lower panel, we arrive at a similar conclusion regarding the conditions under which S-Tonality is particularly helpful for predicting the unemployment trajectory. Here again, when GDP growth is forecast to be subpar, S-Tonality contains substantial predictive power for unemployment, while the unemployment point forecast itself conveys little information. Taken together, these results confirm that the signal from Tonality is indeed most informative when the Greenbook quantitative forecast is calling for subpar growth, a state of affairs that appears to coincide with times of relatively high uncertainty.

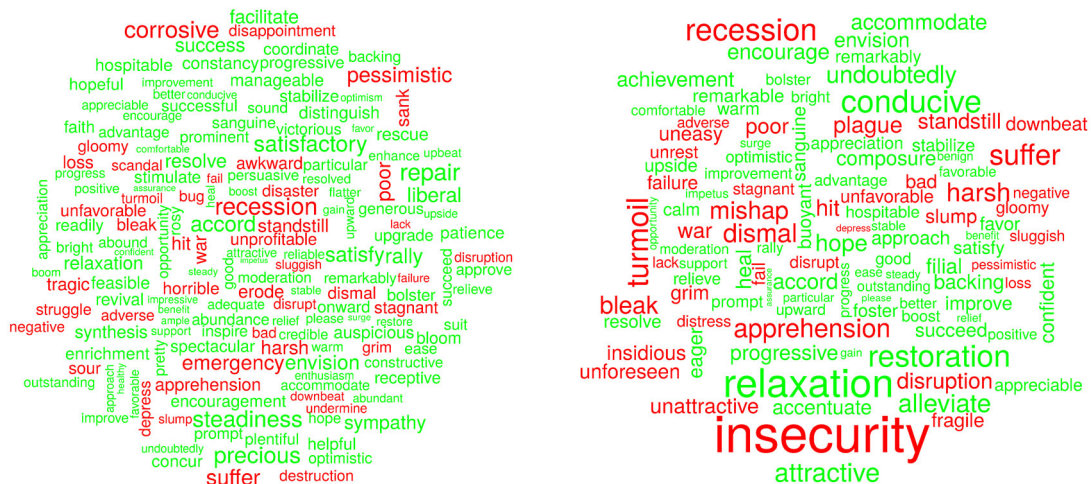
D. A Glance under the Tonality Hood

At this point, we take a closer look at the type of words driving our sentiment measure. A dimension that is relatively quantifiable is whether the signal in Tonality is driven more by variation in negativity or positivity. In particular, S-Tonality is constructed as the net value of two components, which we name S-Positivity (smoothed positive Tonality) and S-Negativity. Somewhat surprisingly, while the correlation between changes in these two components is effectively zero, the correlation between their levels is positive 0.67. This suggests that, on average during periods with higher usage of words conveying positive sentiment, there also tends to be higher usage of words conveying negative sentiment (relative to neutral words). Here, we test whether both components contribute significantly to the forecast by including them separately in forecasting regressions mimicking those in Table 5.

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

We further examine the role played by individual words by creating word clouds in **figure 8** to show the positive and negative words that appear, weighted by their relative importance (unusualness as well as frequency in the corresponding document). Here, we focus on the subset of narratives from Greenbooks where the predictive power of sentiment is strongest, those in which the four-quarter GDP growth forecast was below 2.5 percent (the subpar growth forecast subsample in Table 5 and 6).. The cloud on the left shows the word usage from the Greenbooks for which the GDP *forecast errors* were in the top quartile (again, from the subpar growth forecast subsample), while cloud on the right shows word usage from Greenbooks with GDP forecast errors that were in the bottom quartile. The difference in average Greenbook Tonality between these two subsamples is about 1 standard deviation.

Figure 8: Word cloud on left (right) generated from Greenbooks that precede large positive (negative) GDP surprises



Note: Positive words appear in green and negative words in red. Words are sized in proportion to their weight in the overall Tonality. The word cloud on the left panel from Greenbooks for which GDP forecast errors were in the top quartile while that on the right from Greenbooks for which GDP forecast errors were in the bottom quartile, in both cases from among Greenbooks in which expected four-quarter GDP growth was less than 2.5 percent.

Positive words appear in green and negative words in red. In both clouds, green words are more prevalent than red words, indicating that positive words generally tend to be used much more often. Nonetheless, negative words have relatively greater weight in the cloud to the right (Greenbooks with the most negative forecast errors). Among negative words that are common to the two clouds, such as “recession”, “suffer”, “apprehension”, “turmoil”, and “harsh”, such

words are larger in the right-hand cloud, reflecting their more frequent usage in the Greenbooks preceding negative forecast errors. Also, in this set of Greenbooks), several of the positive words that stand out, such as “relaxation”, “restoration”, and “conducive”, seem to be those associated with improving economic conditions. This is in contrast to the positive-forecast-error world that has a preponderance of positive words but few that stand out from the rest.

IV. The Relevance of Tonality to the Public

So far, our analysis indicates that the information embedded in Tonality appears to contain valuable information for Federal Reserve policymakers, over and above that contained in the staff’s quantitative forecast. In this section, we investigate whether and how the information reflected in Tonality might be of value to market participants outside the Fed. In particular, we examine the information content of Tonality along four dimensions. First, does Tonality complement private-sector economic forecasts in a similar fashion? Second, does Tonality help predict monetary policy? Third, does Tonality predict future stock returns? Finally, we take a brief look at whether the sentiment reflected in Greenbook Tonality shows through to formal FOMC committee public communications.

A. Greenbook Tonality and Blue Chip Forecasts

Tonality of the Greenbook narrative has predictive value for GDP growth and unemployment, conditional on the Greenbook forecast; that is, the narrative can predict errors in point forecasts. Does this reflect some built-in, perhaps conscious, complementarity between the point forecast and the narrative? For instance, does the implied “inefficiency” uniquely affect Greenbook point forecasts? Alternatively, would Tonality contain similar information of value to the public, in that it similarly complements publicly available private-sector economic forecasts? This question can be explored using publicly available forecasts produced around the same time as the Greenbook, in particular, by examining Greenbook Tonality’s marginal predictive power conditional on those forecasts.¹²

¹² This analysis would seem to bear on the issue of whether the Federal Reserve has more information than the median economic forecaster, as in Romer and Romer (2000) and more recently in Nakamura and Steinsson (2018). However, finding that Greenbook Tonality helps to predict forecast errors in, say Blue Chip forecasts does not

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

Table 7 shows the results from regressions analogous to those shown in Table 5, though here Greenbook point forecasts have been replaced by the corresponding consensus Blue Chip (BC) forecasts. The top panel shows regressions predicting four-quarter GDP growth conditional on the Blue Chip forecast and Greenbook Tonality, where we have divided the sample based on whether the Blue Chip consensus calls for four-quarter GDP growth above or below 2.5%. Despite the sample starting a bit later, regression estimates are remarkably similar to those in Table 5. Within the first subsample shown in the first two columns, when expected GDP growth is at least 2.5%, S-Tonality is statistically significant and complements the BC forecast by boosting the regression R-squared from 0.21 to 0.27. As in the regressions that condition on the Greenbook forecast, the BC forecast has practically no predictive power in the below-par growth forecast subsample, but S-Tonality boosts the R-squared from 0.01 to 0.23.

Finally, the bottom panel regressions predicting the trajectory of unemployment again produce similar results. These results, together with the those in the top panel, confirms the view that the economic signal embedded in Tonality is not materially different when used in conjunction with private-sector forecasts than when used in combination with Greenbook forecasts. This raises, though does not answer, the question as to whether the information in Tonality reflects inside knowledge to the Federal Reserve staff, or whether it is also reflected in the thinking of market participants.

B. Tonality as a Predictor of Monetary Policy

Given that Tonality is helpful for predicting economic performance up to four quarters ahead, relative to both internal Fed forecasts as well as private sector forecasts, we consider the

necessarily imply that the Federal Reserve has an information advantage, since some Blue Chip forecasters might also produce narratives along with their point forecasts that convey information similar to that in Tonality.

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

The logic of the hypothesis that Tonality could predict surprises in the Fed funds rate is straightforward, at least for the case of private sector forecasts. To the extent that Blue Chip consensus forecasts of interest rate policy are connected to Blue Chip consensus forecasts for economic growth through something like a “Taylor rule”, then positive economic surprises presaged by Tonality should, in turn, presage positive surprises in the path of policy rates. A key presumption behind this hypothesis is that the effects of such positive economic surprises (or unexpected declines in unemployment rate) are not counterbalanced by downward surprises to inflation, presumably a safe assumption so long as the “Phillips curve” is not positively sloped.¹³

We test the hypothesis that S-Tonality helps to predict monetary policy in Table 8. In particular, we regress realized errors in the Blue Chip consensus forecast of the quarterly average Fed Funds rate on value of S-Tonality at the time of forecast. The first three columns show results for the funds rate forecast at the one quarter, two quarter, and four quarter horizon, respectively. As hypothesized, the coefficient on Tonality is positive and statistically significant at all three horizons; this indicates that higher (lower) Tonality presages policy rates that tend to exceed (fall short of) Blue Chip forecasts.¹⁴

The last three columns add a term spread to the regression, specifically, the difference between the nominal one-year Treasury yield and the federal funds rate at the time of forecast. This can be interpreted as a gauge of market expectations for the short-term interest rate

¹³ The logic for such a connection between the Greenbook forecasts of the federal funds rate and Greenbook forecasts for unemployment seems identical; however, the federal funds “forecast” in the Greenbook has not always been chosen to minimize forecast errors. For instance Reifschneider and Tulip (2017) report that the Greenbook traditionally has taken a more “neutral” approach to the Fed funds rate forecast, that it has tended to “condition on [funds rate] paths that modestly rose or fell over time in a manner that signaled the staff’s assessment ... [of the required] adjustment in policy.” This could result in errors in the funds rate forecast being predictable even when forecast errors in economic performance were not. We therefore consider a test of Tonality’s predictive power for Blue Chip consensus funds rate forecast errors to have a cleaner interpretation.

¹⁴ On the other hand, the intercept in each case is negative, and the intercept values indicate that the funds rate forecast was on average upward biased by 15 basis points per quarter ahead (assuming S-Tonality was zero on average), indicating that forecasters did not anticipate the downward trend of the funds rate over the sample period.

trajectory, though imperfect to the extent there are fluctuations in the one-year term premiums. At all three horizons, the coefficient on the term spread is positive and highly significant; moreover, adding it to the regression lowers the still significant coefficient on S-Tonality at each horizon by as much as half. Thus, it would appear that at least some of the signal for policy rates embedded in Tonality was anticipated by the market.

C. Tonality as a Predictor of Stock Returns

The evidence presented so far indicates the sentiment embedded in the Greenbook narrative contains information about future economic performance that was not incorporated in economists' point forecasts, neither forecasts by Federal Reserve staff nor those by (the consensus of) private forecasters. In addition, Tonality helps to predict errors in Blue Chip consensus forecasts of the monetary policy (the fed funds rate), errors that are directionally consistent with the sign of forecast errors for economic activity. These results beg the question: does Tonality contain information that is not reflected in asset market prices as well? In particular, might Greenbook Tonality also help predict stock market returns? 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 simple extension of that well-trod literature.¹⁵

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. We estimate prediction regressions over the full sample. For the most part, previous studies of news sentiment and stock returns document return

¹⁵ Indeed, given that we already have shown Tonality helps predict some innovations to Fed funds rates, the implications for bond return forecasting seem potentially quite interesting and deserving of careful attention, which we reserve for future study.

predictability only at high frequencies, such as daily returns. More recently, however, Calomiris and Mamayski (2019) find evidence that the sentiment gauged in news articles, aggregated over a month, can help predict stock returns up to one year ahead.

Table 9 shows coefficient estimates from regressions predicting 3-month, 6-month, and 12-month returns on the S&P 500 composite, each in excess of the yield on the maturity-matched Treasury bill. Shown below each specification are 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 S-Tonality. As shown, for all three horizons, the coefficient on S-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 unity—which amounts to roughly 1.5 standard deviations—presages a 3.6 percent higher return over the subsequent 6 months (or 4-meeting period). Although not shown, regressions that also include Tonality Shock, find it has no predictive content, consistent with its irrelevance for economic outcomes.

The adjusted R-squared statistics for the 3-month, 6-month, and 12-month horizons, are 2.1, 4.3 and 5.5, respectively, which are fairly sizable compared with most stock return predictive regressions in the literature for example Welch and Goyal (2008). The out-of-sample R^2 statistics are also positive and nearly as large, in notable contrast with many out-of-sample predictive regressions. If a risk-averse investor were able to take advantage of such information in real time, the gain would be economically meaningful.¹⁶

Given the positive coefficient on Tonality, a natural interpretation for Tonality's predictive value is that it contains information not fully reflected in stock prices at the time

¹⁶ Using the evaluation framework of Campbell and Thompson (2007) for a risk-averse investor suggests this would boost expected 6-month returns by 9.1 percent. In particular, the risky asset return can be expressed as the sum of unconditional expected return on the risky asset (μ), the signal (T_t), and a random shock (e) with mean zero and variance σ_e^2 . Letting $S = (\mu - r_f) / ((\sigma_T^2 + \sigma_e^2))^{1/2}$ represent the Sharpe ratio of the risky asset when no signal is observed, and γ represent relative risk-aversion, then the gain in expected return from observing the signal is equal to $\frac{R^2}{(1-R^2)} \frac{(1+S^2)}{\gamma}$. Using 0.26 as the 6 month Sharpe ratio (S), consistent with the Sharpe ratio on stocks over the 1927-2009 period, we calculate a gain in the expected 6-month return of 9.6 percent.

Greenbook is produced but 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 as well as a decline in risk premiums. On the contrary, it seems unintuitive and implausible to interpret Tonality's effect arising from it being a positively-signed risk premium factor; that would have the odd implication that investors demand a lower risk premium when Greenbook sentiment is more negative. The interpretation that Tonality embeds information that is not reflected in stock prices is consistent with this sentiment not being publicly observable. (Indeed, it is arguable that, at the time, even Fed staff was not fully cognizant of the sentiment embedded in Trend Tonality.)

While S-Tonality seems unlikely to be a proxy for the equity risk premium, it could well be correlated with the risk premium. Indeed, we know that Tonality is highly correlated with economic conditions as well as with forecasts of future conditions, including those that are publicly available (Blue Chip). Thus, it could be instructive to filter out the factors affecting Tonality that are likely to be highly correlated with the equity risk premium. Our approach for separating out such factors is to extract what we call "Residual S-Tonality" using a linear regression of S-Tonality on the factors and then using the regression residual.

Specifically, we control for the projected unemployment rate in the current quarter and change in the unemployment rate over the next two quarters. Current unemployment, for instance, is correlated with S-Tonality and also seems likely to be a proxy for business-cycle-driven variation in the equity risk premium, since risk aversion or perceived risk are arguably linked to employment prospects. Indeed, the perceived-risk interpretation is invoked by Schmidt (2016) as the rationale behind the return predictability he documents for initial unemployment claims (which is highly correlated with Current Unemployment in the Greenbook forecast). Another influence on Tonality that we attempt to "remove" is portion related to the degree of optimism reflected in the Greenbook point forecast. For this purpose, we use the projected two-quarter change in unemployment, though results are quite similar if we instead use the GDP growth forecast.¹⁷

¹⁷ For this exercise, we use the two-quarter, rather than the four-quarter, forecast for unemployment because the two-quarter forecast is available going back to 1970, while using the four-quarter forecast would shorten our sample a

Summing up, Residual S-Tonality is first estimated by regressing S-Tonality on Current Unemployment and the expected two-quarter change in unemployment, both from the Greenbook in which Tonality is measured. In this regression, both coefficient estimates are significant, with that on current unemployment negative and that on the unemployment forecast trajectory positive, producing an R-squared of 0.25. Residual S-Tonality is defined as the residual from this regression. As shown in columns 4, 5 and 6 of Table 9, when excess returns are regressed on Residual S-Tonality, we find it is to have substantial predictive power for stock returns over all three horizons, with more highly significant coefficients than what we observed for plain S-Tonality. The in-sample adjusted R-square statistics are roughly three times higher and out-of-sample R-squares are twice as high as those from the analogous regressions with S-Tonality, bolstering the conclusion that the incremental information in Tonality, if available to investors in real time, would be quite valuable.

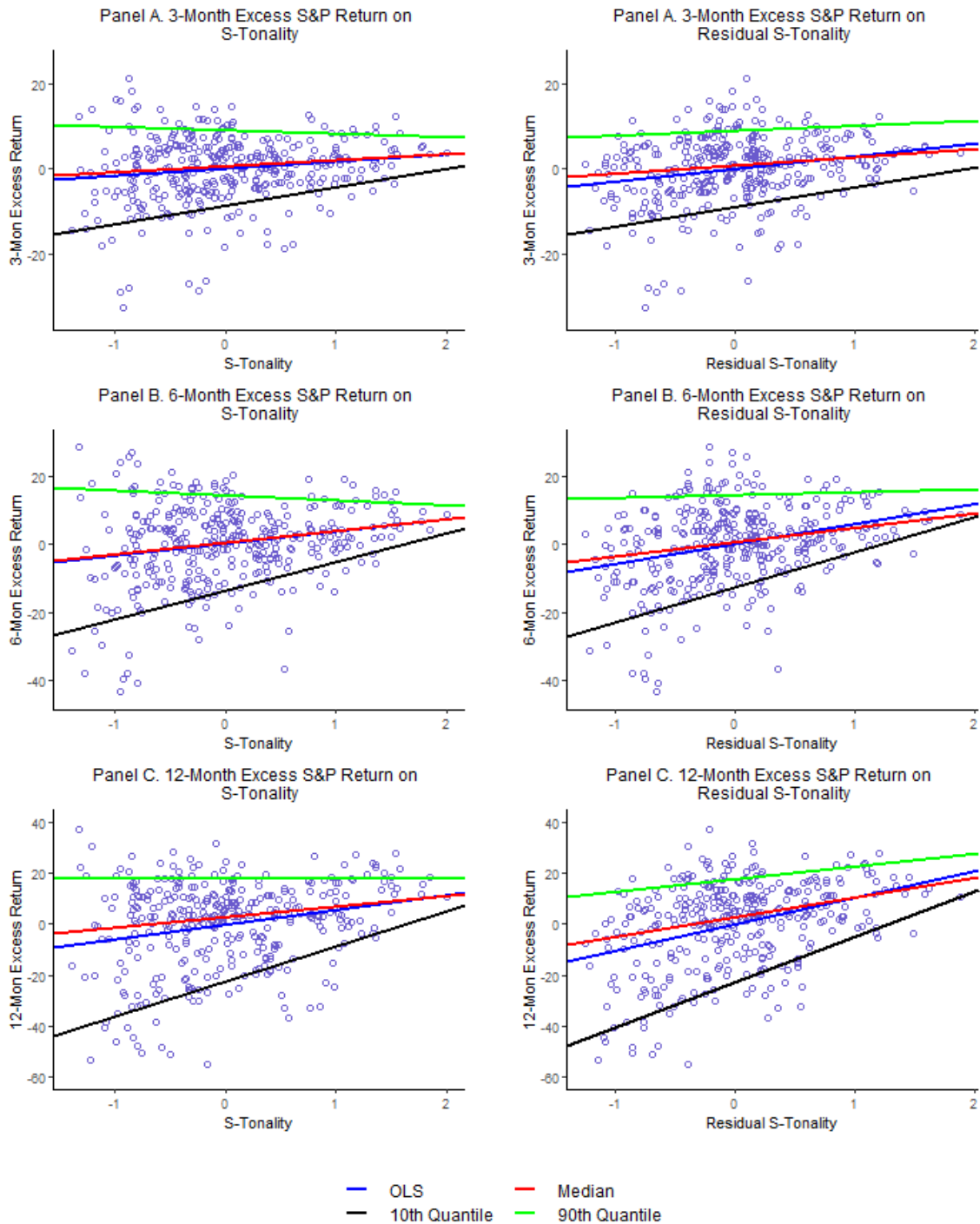
Does the pattern of stock return predictability echo that for the predictability of economic forecast errors? First, we examine whether the predictive power of Tonality for stock returns is associated with downside risk; that is, are is it strongest toward the lower end of potential outcomes—low to negative excess returns—similar to the Tonality’s signal for the distribution of economic forecast errors? **Figure 9** shows quantile regression estimates for 3-month, 6-month and 12-month excess returns, conditioned on S-Tonality in the left panels. At all three horizons, we find Trend Tonality to have its largest predictive effects for returns toward the lower tail of the return distributions, mirroring our findings for macroeconomic predictability. Analogous estimates for residual S-Tonality, shown in the right panels, are qualitatively similar and statistically stronger on balance.

Finally, we examine whether Tonality’s predictive power for returns varies depending on forecasted strength of the economy (which was shown to strongly correlate with macroeconomic uncertainty). In Table 10, we estimate the same regressions as in Table 9, but with subsamples conditioned whether the four-quarter point forecast call for subpar GDP growth ($< 2.5\%$). Comparing results in column 1 to column 3, where excess returns are conditioned on S-Tonality, we do not find greater predictive power in the subsample with sub-par expected growth. On the

few years. Since the two are highly correlated, for the shorter sample, it is pretty much immaterial which horizon is used.

other hand, when returns are conditioned on Residual S-Tonality, a more powerful predictor in the full sample, we find that predictability is much stronger when the forecast calls for subpar growth (column 4) than otherwise (column 2)). Taking the 6-month return horizon for instance, when the forecast calls for subpar growth, the coefficient on Residual S-Tonality is 13.44 and the adjusted R-squared 19 percent, compared to a coefficient of 3.76 and an adjusted R-squared of 5 percent in subsample with stronger growth forecasts.

Figure 9: Regressions Predicting Forecast Errors: OLS, 10th and 90th Quantiles



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

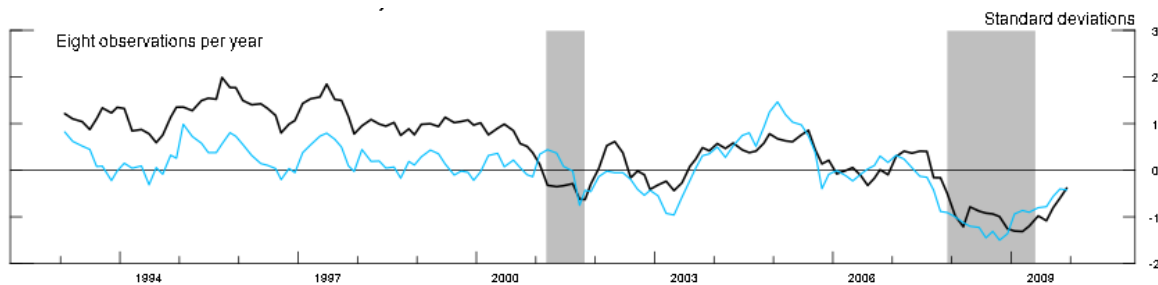
D. Is Greenbook Tonality Communicated to the Public?

To gauge the extent to which the sentiment of the Greenbook narrative is transmitted to the public through FOMC communications, we measure the Tonality of the two regular communications issued to the public, (i) FOMC statements and (ii) minutes of the FOMC meetings. In February 1993, the committee began issuing minutes of its deliberations after a delay of several weeks but prior to the subsequent meeting. In February 1994, the FOMC committee began releasing relatively terse statements explaining its actions or stance, at first sporadically and then after every meeting starting May 1999.

For each set of communications, Tonality is measured by counting positive and negative word usage in those documents and normalizing using the analogous tdf-if routine used in our analysis of the Greenbooks. The resultant time series for statement Tonality is uncorrelated with Greenbook Tonality (0.04 for full sample, same as the post-May 1999 sample). In contrast, the correlation of 0.50 between Minutes Tonality and Greenbook Tonality would appear to be quite substantial. Constructing Trend Minutes Tonality, we find its correlation with the analogous Trend Tonality for Greenbook to be even higher, at 0.74. As shown in **figure 10**, those two measure of sentiment look quite similar, and would appear even more so if not for their divergent trends in early 2001.

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

Figure 10: Minutes versus Greenbook Trend Tonality



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

V. Summary, Interpretation, and Conclusions

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

The finding that Tonality predicts errors in Blue Chip funds rate forecasts indicates that Tonality conveys policy-relevant information. The finding that Tonality predicts future stock returns, while notable in its own right, seems not entirely surprising once we have established its

ability to predict unexpected economic growth. Given that lower Tonality predicts both greater downside economic risks as well as much lower-than-average returns, the time varying return documented here would not seem to reflect expected compensation for expected risk. Rather, these results suggest that equity prices do not contemporaneously impound all the information about the potential evolution of the economy that is impounded in the forecast narrative.

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

VIII. References

- Adams, Partick, Tobais Adrian, Nina Boyarchenko, and Domenico Giannone. 2020. *Forecasting Macroeconomic Risks*. Discussion Paper, London: Centre for Economic Policy Research
- Adrian, Tobias, Nina Boyarchenko, and Domenico Giannone. 2019. "Vulnerable growth." *American Economic Review* 109 (4): 1263-89.
- Aggarwal, Raj, Sunil Mohanty, and Frank Song. 1995. "Are Survey Forecasts of Macroeconomic Variables Rational?" *The Journal of Business* 68 (1): 99-119.

- 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. 2019. "How news and its context drive risk and returns around the world." *Journal of Financial Economics* 133 (2): 299-336.
- Campbell, John Y. 1987. "Stock returns and the term structure." *Journal of Financial Economics* 18 (2): 373-399.
- Campbell, John Y., and Samuel B. Thompson. 2007. "Predicting excess stock returns out of sample: Can anything beat the historical average?" *The Review of Financial Studies* (21): 1509-1531.
- Carvalho, Carlos, Eric Hsu, and Fernanda Nechio. 2016. "Measuring the effect of the zero lower bound on monetary policy." *Federal Reserve Bank of San Francisco Working Paper* 1-32.
- Clark, Todd E., Michael W. McCracken, and Elmar and Mertens. 2020. "Modeling time-varying uncertainty of multiple-horizon forecast errors." *Review of Economics and Statistics* 102 (1): 17-33.
- Coibion, Olivier, and Yuriy and Gorodnichenko. 2015. "Information rigidity and the expectations formation process: A simple framework and new facts." *American Economic Review* 105 (8): 2644-78.
- Coibion, Olivier, and Yuriy Gorodnichenko. 2012. "What can survey forecasts tell us about information rigidities?" *Journal of Political Economy* 120 (1): 116-159.
- Croushore, Dean, and Tom Stark. 2001. "A real-time data set for macroeconomists." *Journal of Econometrics* 105 (1): 111-130.
- Da, Zhi, Joseph Engelberg, and Pengjie Gao. 2014. "The sum of all FEARS." *The Review of Financial Studies* 28 (1): 1-32.
- D'Agostino, Antonello, and Karl Whelan. 2008. "Federal Reserve Information during the Great Moderation." *Journal of the European Economic Association* 6 (2-3): 609-620.

- Das, Sanjiv R, and Mike Y Chen. 2007. "Yahoo! for Amazon: Sentiment extraction from small talk on the web." *Management Science* 53 (9): 1375--1388.
- Dovern, Jonas, Ulrich Fritsche, Prakash Loungani, and Natalia and Tamirisa. 2015. "Information rigidities: Comparing average and individual forecasts for a large international panel." *International Journal of Forecasting* 31: 144-154.
- Feng, Xingdong, Xuming He, and Jianhua Hu. 2011. "Wild bootstrap for quantile regression." *Biometrika* 98 (4): 995-999.
- Garcia, Diego. 2013. "Sentiment during recessions." *Journal of Finance* 68 (3): 1267-1300.
- Gürkaynak, Refet S., Brian P. Sack, and Eric T. Swanson. 2007. "Market-based measures of monetary policy expectations." *Journal of Business & Economic Statistics* 25 (2): 201-212.
- Gürkaynak, Refet S., Brian Sack, and Eric Swanson. 2005. "The sensitivity of long-term interest rates to economic news: evidence and implications for macroeconomic models." *The American Economic Review* 95 (1): 425-436.
- Hansen, Stephen, and Michael McMahon. 2016. "Shocking language: Understanding the macroeconomic effects of central bank communication." *International Economic Review* 99: S114-S133.
- Hansen, Stephen, Michael McMahon, and and Andrea Prat. 2017. "Transparency and deliberation within the FOMC: a computational linguistics approach." *The Quarterly Journal of Economics* 133 (2): 801-870.
- Hengge, Martina. 2019. *Uncertainty as a predictor of economic activity*. Working Paper, Economics Section, The Graduate Institute of International Studies.
- Heston, Steven L., and Nitish Ranjan Sinha. 2017. "News vs. Sentiment: Predicting Stock Returns from News Stories." *Financial Analyst Journal* 73: 67-83.
- Jones, Jacob, T., Tara, M. Sinclair, and O. Herman Stekler. 2019. "A textual analysis of Bank of England growth forecasts." *International Journal of Forecasting* 1-10.
- Jurado, Kyle, Sydney Ludvigson, and Serena Ng. 2015. "Measuring Uncertainty." *American Economic Review* 105 (3): 1177-1216.
- Liu, Regina. 1988. "Bootstrap procedures under some non-iid models." *The Annals of Statistics* 16 (4): 1696-1708.
- Loughran, Tim, and and Bill McDonald. 2011. "When a liability is not a liability? Textual analysis, dictionaries, and 10-Ks." *The Journal of Finance* 66 (1): 35-65.

- Nakamura, Emi, and Jón Steinsson. 2018. "High-frequency identification of monetary non-neutrality: the information effect." *The Quarterly Journal of Economics* 133 (3): 1283-1330.
- Newey, Whitney K., and Kenneth D. West. 1994. "Automatic lag selection in covariance matrix estimation." *The Review of Economic Studies* 61 (4): 631-653.
- Nordhaus, William D. 1987. "Forecasting Efficiency: Concepts and Applications." *Review of Economics and Statistics* 667-674.
- Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, et al. 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.
- Rogers, John H., and Xu Jiawen. 2019. "How Well Does Economic Uncertainty Forecast Economic Activity?" *FEDS Working Paper* (2019-085).
doi:<https://doi.org/10.17016/FEDS.2019.085>.
- Romer, Christina D., and David H. Romer. 2000. "Federal Reserve information and the behavior of interest rates." *The American Economic Review* 90: 429-457.
- Rudebusch, Glenn D., and John C. Williams. 2009. "Forecasting recessions: the puzzle of the enduring power of the yield curve." *Journal of Business & Economic Statistics* 27 (4): 492-503.
- Schmeling, Maik and Wagner, Christian. 2017. *Does central bank tone move asset prices?* SSRN Working Paper, 1-75.
- Schmidt, Lawrence D. W. 2016. *Climbing and Falling Off the Ladder: Asset Pricing Implications of Labor Market Event Risk*. SSRN Working Paper, University of Chicago.
- Shiller, Robert, T. 2017. "Narrative Economics." *American Economic Review* 107 (4): 967-1004.
- Sinclair, Tara M., Fred Joutz, and Herman O. Stekler. 2010. "Can the Fed predict the state of the economy?" *Economics Letters* 108 (1): 28-32.
- Smirnov, Sergey V., and Avdeeva Daria. 2016. "Wishful Bias in Predicting US Recessions: Indirect Evidence." *Higher School of Economics Research Paper*.
doi:<http://dx.doi.org/10.2139/ssrn.2781923> .

- Stock, James H., and Mark W. Waston. 2003. "Forecasting Output and Inflation: The Role of Asset Prices." *Journal of Economic Literature* 788-829.
- Stock, James H., and Mark W. Watson. 2007. "Why has US inflation become harder to forecast?" *Journal of Money, Credit and Banking* 39 (s1): 3-33.
- Tetlock, Paul C. 2007. "Giving content to investor sentiment: The role of media in the stock market." *The Journal of Finance* 62 (3): 1139-1168.
- Welch, Ivo, and Amit Goyal. 2008. "A Comprehensive Look at The Empirical Performance of Equity Premium Prediction." *The Review of Financial Studies* 21 (4): 1455-1508.
- Wu, Chien-Fu Jeff. 1986. "Jackknife, bootstrap and other resampling methods in regression analysis." *The Annals of Statistics* 14 (4): 1261-1295.
- Zarnowitz, V., and P. A. Braun. 1993. "Twenty-Two Years of the NBER-ASA Quarterly Economic Outlook Surveys: Aspects and Comparisons of Forecasting Performance." In *Business Cycles, Indicators, and Forecasting*, by J., H. Stock and M. W. and Watson. Chicago: University of Chicago Press .
- Zarnowitz, Victor. 1985. "Rational Expectations and macroeconomic forecasts." *Journal of Business & Economic Statistics* 3 (4): 293-311.
- Zheng Tracy Ke, Bryan T. Kelly and Dacheng Xiu. 2019. *Predicting Returns with Text Data*. Cambridge, MA: NBER Working Paper Series.

Table 1: Correlation between Greenbook text sentiment measures, Greenbook forecast variables and other information measures

	Tonality	S-Tonality	Tonality Shock
4-qtr Δ GDP	0.22***	0.29***	0.04
4-qtr Δ GDP revision	0.26***	0.20***	0.20***
4-qtr Δ Unemp	-0.33***	-0.46***	-0.03
4-qtr Δ Unemp revision	-0.27***	-0.25***	-0.16***
4-qtr Δ Inflation	-0.33***	-0.49***	0.01
4-qtr Δ Inflation revision	-0.10*	-0.13**	-0.02
Current-qtr Unemployment	-0.07	-0.24***	0.15***
EPU-Gbk	-0.16***	-0.14***	-0.11**
EPU-BBD	-0.12**	-0.13**	-0.06
MACROU	-0.47***	-0.68***	-0.01
NFCI	-0.47***	-0.67***	-0.01
Gbk stock return	0.22***	0.15***	0.19***

Notes: Correlation coefficients are computed pairwise. The top panel shows correlation between Greenbook text measures and Greenbook forecast variables. Tonality is the weighted difference between positive and negative terms in the Greenbook. Tonality shock is the deviation of Tonality from S-Tonality, its smoothed version. The bottom panel shows the correlation between Greenbook text measures and other information measures matched to the corresponding Greenbook. Tonality shock is 4-qtr Δ GDP is the cumulative 4-qtr GDP growth forecast from the publication of the Greenbook. GDP Revision is the revision to the 4-Qtr GDP growth since the previous Greenbook. Unemployment and Inflation forecast and revision are similarly defined with respect to the change in the unemployment rate and cumulative change in inflation. Of the non-forecast terms, EPU-Gbk is the scaled count of "uncertain" and "uncertainty" in the Greenbook texts, while EPU-BBD is the Economic Policy Uncertainty index (Baker, Bloom and Davis 2016). MACROU and NFCI are short for the 12-month Macroeconomic Uncertainty measure developed by Jurado, Ludvigson and Ng (2015) and the Federal Reserve Bank of Chicago's National Financial Conditions Index respectively. The final row is the cumulative stock return between Greenbook internal publication dates; roughly a week prior to the FOMC meetings. *p<0.1; **p<0.05; ***p<0.01

Table 2: Regressions Predictiong Forecast Errors with Tonality or S-Tonality OLS and Quantile (10,90) Regression Statistics

Forecast Error	Regressor	OLS			10th Percentile			90th Percentile		
		Coef.	p-val	R2	Coef.	p-val	R2	Coef.	p-val	R2
GDP 1-qtr	Tonality	0.16	0.03	0.03	0.31	0.00	0.08	0.08	0.26	0.00
GDP 1-qtr	S-Tonality	0.22	0.11	0.02	0.57	0.00	0.11	-0.13	0.36	0.00
GDP 4-qtr	Tonality	0.45	0.03	0.04	0.75	0.00	0.07	0.21	0.30	0.00
GDP 4-qtr	S-Tonality	0.80	0.06	0.07	1.61	0.00	0.12	-0.19	0.38	0.00
Unemp 1-qtr	Tonality	-0.05	0.06	0.02	0.06	0.03	0.00	-0.16	0.00	0.05
Unemp 1-qtr	S-Tonality	-0.09	0.09	0.02	0.12	0.00	0.02	-0.23	0.00	0.10
Unemp 4-qtr	Tonality	-0.23	0.01	0.07	0.00	1.00	0.00	-0.47	0.00	0.12
Unemp 4-qtr	S-Tonality	-0.43	0.01	0.12	0.06	0.41	0.00	-0.84	0.00	0.21
CPI 1-qtr	Tonality	-0.01	0.74	-0.00	0.12	0.00	0.05	-0.15	0.00	0.04
CPI 1-qtr	S-Tonality	0.01	0.94	-0.00	0.20	0.00	0.08	-0.30	0.00	0.07
CPI 4-qtr	Tonality	-0.02	0.93	-0.00	0.55	0.00	0.07	-0.49	0.06	0.02
CPI 4-qtr	S-Tonality	0.07	0.84	-0.00	0.94	0.00	0.14	-1.04	0.00	0.05

Notes: The estimates are from univariate regressions of forecast error on Tonality and S-Tonality from January 1972 through December 2009. Columns 1, 4 and 7 report the Tonality and S-Tonality coefficients from OLS and quantile regressions. p-values from OLS regressions in column 2 are computed using standard errors corrected for autocorrelation following Newey West 1994 using $(2*k + 1)$ lags for k quarter out forecast error. For quantile regressions, p-values in columns 5 and 8 follow from wild-bootstrap standard errors developed by Feng, He and Hu (2011) with 1000 resamples. The Psuedo R^2 statistics in columns 6 and 9 are as described in Koenker and Machado (1999). *p<0.1; **p<0.05; ***p<0.01

Table 3: Marginal Effects of S-Tonality on Forecast Errors with Control Variables (Four quarter forecasts)

	OLS			10th Percentile			90th Percentile		
	Coef.	p-val	R2	Coef.	p-val	R2	Coef.	p-val	R2
GDP growth forecast error									
S-Tonality (No Control)	0.80	0.06	0.07	1.61	0.00	0.12	-0.19	0.37	0.00
Ctrl for MACROU	1.35	0.00	0.10	0.84	0.00	0.20	1.19	0.00	0.16
Ctrl for NFCI	1.06	0.01	0.08	1.07	0.00	0.18	0.88	0.00	0.09
Ctrl for Stock Return	0.72	0.08	0.09	1.35	0.00	0.15	-0.35	0.07	0.02
Ctrl for Revision	0.73	0.08	0.06	1.42	0.00	0.11	-0.30	0.13	0.02
Unemployment forecast error									
S-Tonality (No Control)	-0.43	0.01	0.12	0.06	0.42	0.00	-0.84	0.00	0.21
Ctrl for MACROU	-0.34	0.06	0.12	-0.29	0.05	0.05	-0.44	0.00	0.25
Ctrl for NFCI	-0.34	0.10	0.12	-0.04	0.70	0.01	-0.67	0.00	0.22
Ctrl for Stock Return	-0.39	0.01	0.16	0.03	0.75	0.01	-0.73	0.00	0.23
Ctrl for Revision	-0.40	0.01	0.11	0.10	0.13	0.00	-0.82	0.00	0.20
CPI Inflation forecast error									
S-Tonality (No Control)	0.07	0.84	-0.00	0.94	0.00	0.14	-1.04	0.00	0.05
Ctrl for MACROU	-0.52	0.15	0.06	0.20	0.44	0.21	-0.95	0.00	0.05
Ctrl for NFCI	-0.19	0.55	0.01	0.27	0.30	0.19	-0.99	0.00	0.06
Ctrl for Stock Return	0.08	0.81	-0.00	0.94	0.00	0.14	-1.02	0.00	0.05
Ctrl for Revision	0.13	0.70	-0.00	0.94	0.00	0.15	-0.92	0.00	0.05

Notes: The estimates are from regressions of forecast error on Tonality and S-Tonality from January 1972 through December 2009. The controls Macroeconomic Uncertainty (12 month MACROU from Jurado, Ludvigson and Ng (2015)), the Federal Reserve Bank of Chicago's National Financial Conditions Index (NFCI), the market return between green books (Stock Return), and the change in forecast from the prior greenbook (Revision). Columns 1, 4 and 7 report the Tonality and S-Tonality coefficients from OLS and quantile regressions. p-values from OLS regressions in column 2 are computed using standard errors corrected for autocorrelation following Newey West 1994 using $(2*k + 1)$ lags for k quarter out forecast error. For quantile regressions, p-values in columns 5 and 8 follow from wild-bootstrap standard errors (Feng, He and Hu 2011) with 1000 resamples. The Psuedo R^2 statistics in columns 6 and 9 are as described in (Koenker and Machado 1999). *p<0.1; **p<0.05; ***p<0.01

Table 4: OLS Regressions of Realized Performance on Forecast and S-Tonality Conditioned on Uncertainty

Dependent Variable: Realized four-quarter GDP growth				
	Normal/Low Uncertainty		High Uncertainty	
Forecast	0.71*** (0.06)	0.80*** (0.06)	0.71*** (0.14)	0.17 (0.17)
S-Tonality		0.84*** (0.15)		4.81*** (1.15)
Intercept	1.10*** (0.28)	0.47* (0.28)	0.07 (0.36)	4.14*** (0.95)
Observations	240	240	84	84
Adjusted R ²	0.40	0.47	0.20	0.37
Residual Std. Error	1.43	1.35	3.27	2.92

Dependent Variable: Realized four-quarter change in Unemployment				
	Normal/Low Uncertainty		High Uncertainty	
Forecast	1.06*** (0.06)	1.07*** (0.06)	0.94*** (0.17)	0.27 (0.17)
S-Tonality		-0.17*** (0.04)		-2.43*** (0.51)
Intercept	-0.18*** (0.03)	-0.13*** (0.04)	0.41 (0.25)	-0.56* (0.28)
Observations	240	240	84	84
Adjusted R ²	0.58	0.60	0.25	0.48
Residual Std. Error	0.51	0.50	1.44	1.19

Notes. The estimates are from regressions of realizations on the Staff Forecast and S-Tonality using data spanning January 1972 through December 2009. Samples are split into Normal/Low (High) Uncertainty by being below (above) the 75th percentile of the 12-month Macro Uncertainty developed by Jurado, Ludvigson and Ng (2015). Standard errors are computed using a wild-bootstrap with 1000 resamples developed by Wu (1986) and Liu (1988). *p<0.1; **p<0.05; ***p<0.01

Table 5: OLS Regressions of Realized Performance on Forecast and S-Tonality Conditioned on GDP forecast

Dependent Variable: Realized four-quarter GDP growth				
	$\mathbf{E}(GDP) \geq 2.5$		$\mathbf{E}(GDP) < 2.5$	
Forecast	0.77*** (0.07)	0.88*** (0.08)	0.22 (0.14)	-0.25 (0.15)
S-Tonality		1.01*** (0.25)		2.26*** (0.45)
Intercept	0.87** (0.35)	0.14 (0.44)	0.49** (0.25)	1.99*** (0.31)
Observations	239	239	85	85
Adjusted R ²	0.31	0.39	0.005	0.22
Residual Std. Error	1.78	1.67	2.60	2.31
Dependent Variable: Realized four-quarter change in Unemployment				
	$\mathbf{E}(GDP) \geq 2.5$		$\mathbf{E}(GDP) < 2.5$	
Forecast	1.07*** (0.10)	1.06*** (0.09)	0.53*** (0.17)	-0.06 (0.18)
S-Tonality		-0.19*** (0.06)		-1.27*** (0.20)
Intercept	-0.21*** (0.05)	-0.17*** (0.05)	0.98*** (0.26)	1.03*** (0.20)
Observations	239	239	85	85
Adjusted R ²	0.49	0.50	0.08	0.35
Residual Std. Error	0.60	0.59	1.24	1.04

Notes. The estimates are from regressions of realizations on the Staff Forecast and S-Tonality using data spanning January 1972 through December 2009. Samples are split into Strong/Normal (Subpar) Growth samples if above (below) 2.5 percent; roughly the 25th percentile of the 4-qtr Cumulative GDP Forecast. Standard errors are computed using a wild-bootstrap with 1000 resamples. *p<0.1; **p<0.05; ***p<0.01

Table 6: OLS Regressions of Realized Performance on Forecast and S-Tonality Subcomponents

Dependent Variable: Realized four-quarter GDP growth				
	$\mathbf{E}(GDP) \geq 2.5$		$\mathbf{E}(GDP) < 2.5$	
Forecast	0.88*** (0.08)	0.90*** (0.09)	-0.25 (0.15)	-0.26* (0.16)
S-Tonality	1.01*** (0.25)		2.26*** (0.45)	
S-Positivity		0.98*** (0.23)		2.27*** (0.46)
S-Negativity		-0.74*** (0.25)		-2.05*** (0.61)
Intercept	0.14 (0.44)	0.05 (0.47)	1.99*** (0.31)	1.99*** (0.31)
Observations	239	239	85	85
Adjusted R ²	0.39	0.39	0.22	0.21
Residual Std. Error	1.67	1.67	2.31	2.32
Dependent Variable: Realized four-quarter change in Unemployment				
	$\mathbf{E}(GDP) \geq 2.5$		$\mathbf{E}(GDP) < 2.5$	
Forecast	1.06*** (0.09)	0.93*** (0.11)	-0.06 (0.18)	0.06 (0.17)
S-Tonality	-0.19*** (0.06)		-1.27*** (0.20)	
S-Positivity		-0.26*** (0.06)		-1.21*** (0.19)
S-Negativity		0.61*** (0.11)		1.84*** (0.28)
Intercept	-0.17*** (0.05)	-0.14*** (0.05)	1.03*** (0.20)	0.91*** (0.19)
Observations	239	239	85	85
Adjusted R ²	0.50	0.53	0.35	0.41
Residual Std. Error	0.59	0.57	1.04	0.99

Notes. The estimates are from regressions of realizations on the Staff Forecast and S-Tonality using data spanning January 1972 through December 2009. Samples are split into Strong/Normal (Subpar) Growth samples if above (below) 2.5 percent; roughly the 25th percentile of the 4-qtr Cumulative GDP Forecast. S-Positivity and S-Negativity are the components whose net value is used to form S-Tonality. Standard errors are computed using a wild-bootstrap with 1000 resamples. *p<0.1; **p<0.05; ***p<0.01

Table 7: OLS Regressions of Realized Performance on Blue Chip Forecast and S-Tonality

Dependent Variable: Realized four-quarter GDP growth				
	$\mathbf{E}(GDP) \geq 2.5$		$\mathbf{E}(GDP) < 2.5$	
BC Forecast	0.84*** (0.15)	0.97*** (0.15)	0.37 (0.24)	-0.03 (0.24)
S-Tonality		0.62*** (0.15)		1.85*** (0.38)
Intercept	0.86 (0.56)	0.13 (0.56)	0.09 (0.57)	1.61** (0.66)
Observations	178	178	60	60
Adjusted R ²	0.21	0.27	0.01	0.23
Residual Std. Error	1.37	1.33	2.53	2.22
Dependent Variable: Realized four-quarter change in Unemployment				
	$\mathbf{E}(GDP) \geq 2.5$		$\mathbf{E}(GDP) < 2.5$	
BC Forecast	1.45*** (0.11)	1.48*** (0.11)	0.82*** (0.24)	0.22 (0.22)
S-Tonality		-0.27*** (0.06)		-1.16*** (0.17)
Intercept	-0.16*** (0.05)	-0.04 (0.06)	0.98*** (0.21)	0.92*** (0.15)
Observations	178	178	60	60
Adjusted R ²	0.53	0.57	0.11	0.38
Residual Std. Error	0.56	0.53	1.32	1.10

Notes. The estimates are from regressions of realizations on the Blue Chip Consensus Forecast and S-Tonality using data spanning July 1980 through December 2009. Samples are split into Strong/Normal (Subpar) Growth samples if above(below) 2.5 percent; roughly the 25th percentile of the 4-qtr Cumulative GDP Forecast. Standard errors are computed using a wild-bootstrap with 1000 resamples. *p<0.1; **p<0.05; ***p<0.01

Table 8: OLS Regressions of Blue Chip Fed Funds Forecast Errors on S-Tonality

	1-Qtr	2-Qtr	4-Qtr	1-Qtr	2-Qtr	4-Qtr
S-Tonality	0.13*** (0.05)	0.26*** (0.10)	0.60*** (0.22)	0.07* (0.04)	0.16* (0.09)	0.43** (0.19)
Term Spread				0.32*** (0.08)	0.57*** (0.19)	0.99** (0.44)
Intercept	-0.15*** (0.05)	-0.32*** (0.12)	-0.75*** (0.28)	-0.19*** (0.05)	-0.40*** (0.11)	-0.88*** (0.26)
Observations	192	192	192	192	192	192
Adjusted R ²	0.05	0.06	0.09	0.18	0.16	0.18
Residual Std. Error	0.37	0.71	1.30	0.34	0.67	1.23

Notes. The estimates are from regressions of Blue Chip forecast errors on S-Tonality from February 1986 through December 2009. The Term Spread is defined here as the difference between a 1-Year Treasury Note and the Fed Funds rate when the Greenbook is produced. Standard errors are corrected for autocorrelation following Newey West 1994 using $(2*k + 1)$ lags for k quarter out forecast error. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table 9: Regressions Predicting Excess S&P 500 Returns

	1-Qtr	2-Qtr	4-Qtr	1-Qtr	2-Qtr	4-Qtr
S-Tonality	1.718*	3.571**	5.775*			
	(0.906)	(1.742)	(2.950)			
Residual S-Tonality				2.991***	5.927***	10.490***
				(0.875)	(1.685)	(3.080)
Intercept	0.055	0.094	-0.113	0.084	0.154	-0.017
	(0.621)	(1.160)	(2.172)	(0.596)	(1.094)	(2.016)
Out-of-sample R^2	0.021	0.024	0.047	0.044	0.075	0.066
Observations	357	357	357	357	357	357
Adjusted R^2	0.021	0.043	0.055	0.050	0.090	0.138
Residual Std. Error	7.799	11.544	16.453	7.681	11.255	15.709

Notes. Results are from regressions from January 1972 through December 2009. Residual S-Tonality is the residuals from an OLS regression of S-Tonality on the current quarter unemployment forecast and 2-quarter change in unemployment forecast; the adjusted R-square of the regression is 0.254. Standard errors are corrected for autocorrelation following Newey West 1994 using $(2*k + 1)$ lags for k quarter out forecast error. The Out-of-sample R^2 are calculated over the period that begins 64 meetings into the start of the sample through December 2009. *p<0.1; **p<0.05; ***p<0.01

Table 10: Regressions Predicting Excess S&P 500 Return Conditioned on GDP forecast

Dependent Variable: 3-Month Excess S&P 500 Returns				
	$\mathbf{E}(GDP) \geq 2.5$		$\mathbf{E}(GDP) < 2.5$	
S-Tonality	1.11*		2.29	
	(0.66)		(1.50)	
Residual S-Tonality		1.51**		8.12***
		(0.68)		(1.98)
Intercept	0.42	0.52	0.50	0.63
	(0.52)	(0.46)	(1.03)	(1.01)
Observations	239	239	85	85
Adjusted R ²	0.01	0.01	0.01	0.16
Residual Std. Error	6.67	6.64	10.19	9.41
Dependent Variable: 6-Month Excess S&P 500 Returns				
	$\mathbf{E}(GDP) \geq 2.5$		$\mathbf{E}(GDP) < 2.5$	
S-Tonality	2.65**		3.73*	
	(1.01)		(2.15)	
Residual S-Tonality		3.76***		13.44***
		(0.98)		(2.97)
Intercept	0.54	0.76	0.19	0.44
	(0.75)	(0.66)	(1.39)	(1.50)
Observations	239	239	85	85
Adjusted R ²	0.03	0.05	0.01	0.19
Residual Std. Error	9.57	9.45	15.51	14.08
Dependent Variable: 12-Month Excess S&P 500 Returns				
	$\mathbf{E}(GDP) \geq 2.5$		$\mathbf{E}(GDP) < 2.5$	
S-Tonality	5.31***		6.02**	
	(1.56)		(2.99)	
Residual S-Tonality		7.27***		22.63***
		(1.55)		(3.87)
Intercept	-0.03	0.44	0.04	0.59
	(1.12)	(1.00)	(2.04)	(1.82)
Observations	239	239	85	85
Adjusted R ²	0.05	0.08	0.02	0.28
Residual Std. Error	14.61	14.32	21.34	18.28

Notes. Results are from regressions of realizations on the Staff Forecast and S-Tonality using data spanning January 1972 through December 2009. Residual S-Tonality is the residuals from an OLS regression of S-Tonality on the current quarter unemployment forecast and 2-quarter change in unemployment forecast; the adjusted R-square of the regression is 0.254. Samples are split into Strong (Weak) Growth samples if above (below) the 75th percentile of the 4-qtr Cumulative GDP Forecast. Standard errors are computed using a wild-bootstrap with 1000 resamples. *p<0.1; **p<0.05; ***p<0.01

Appendix A: Text analysis

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

List of 231 positive words

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

List of 102 negative words

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

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

To provide some sense of the words that go into the constitution of Tonality, we provide word clouds showing the 50 most prominent positive and negative words in Greenbook during a couple different time periods. **Figure A1** shows two side-by-side word clouds for the 50 most prominent positive words in Greenbooks during two periods, 1994-1998 and 2005-2009. Word size is proportional to its contribution to Tonality, that is, its contribution to the sum of tf-idf weights during the five-year window. Overall, the positive word cloud is a bit bigger during the later period. The substantial overlap in influential words during these two periods suggests little language drift, whereby many words fall out of favor and are replaced by new ones. The most important positive word in both periods is “upward”, followed closely by “positive.” On the other hand, the words “favorable” and “moderation” are more prominent during 1994-1998.

Figure A2 shows two side-by-side word clouds for the 50 most prominent negative words in Greenbooks during the same two periods. The most prominent negative word in both samples is “negative”, followed by “sluggish.” Overall, negative words are more prominent in the later period as indicated by the larger word sizes in that cloud. For example, the words “adverse” and “sluggish” are more prominent in 2005-2009 period.

Figure A1: Word cloud for fifty most positive words in the Greenbook.



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

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



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

Appendix B: Data

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

1. Tonality Variables

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

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

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

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

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

2. Economic Variables

Historical realized values

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

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

$$Act_RGDP_growth_i = 100 * ((RGDP_i / RGDP_{-1}) - 1)$$

Similarly, the unemployment change, we use the quarter prior to the Greenbook publication as base value. $Act_Unemployment_{-1}$ is the value of $Unemployment$ from the previous quarter and $Unemployment_i$ is the

value of *Unemployment* *i* quarters into the future. We then compute the *i* quarters ahead unemployment change as following:

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

Growth in CPI is instead calculated using the contemporaneous CPI. *Act_CPI₀* 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_RDGP₀ is the staff's projection of the growth from the previous quarter to the current quarter of RGDP. *Staff_RDGP_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_CPI₀ is the staff's projection for the change in CPI from the previous quarter to the current quarter. *Staff_CPI_i* is equal to the projected Q/Q growth *i* quarters into the future. We then compute the *i* quarters ahead cumulative CPI growth as following:

$$Staff_CPI_growth_i = \prod_{k=1}^i Staff_CPI_k$$

Blue Chip Forecasts

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

$$BC_RGDP_growth_i = \prod_{k=0}^i BC_RGDP_k$$

$$BC_Unemployment_change_i = BC_Unemployment_i - BC_Unemployment_{-1}$$

$$BC_CPI_growth_i = \prod_{k=1}^i BC_CPI_k$$

3. Federal Fund Rate Variables

Actuals

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

$Act_FedFunds_{-1}$ is equal to the average Fed funds rate in the previous quarter. $Act_FedFunds_i$ is the average rate i quarters into the future. We define the change in Fed funds rate as follows:

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

Blue Chip Forecast

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

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

4. Revisions

We create revision variables for both the Staff and Blue Chip forecasts. Revisions are defined as the difference between the current forecast and the previous forecast for the same period. In the case that the Greenbook release date is in the first month of the quarter, the forecast from the period before will use one additional forecast period in order to maintain the quarterly alignment. For example, in January the revision for a 1-quarter ahead forecast will be calculated as the current 1-quarter ahead forecast minus the December meeting's 2-quarter ahead forecast. We define the revision for the i quarter ahead projection at meeting t as follows:

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

5. Asset Price Variables

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

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

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

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